

Response to reviewer VWG5

We appreciate the referee's recognition of the value of our paper, and insightful comments helpful to improve our work. In the following, we address the proposed weakness and questions.

The paper lacks a rigorous theoretical explanation for why rank space leads to improved signal-to-noise ratio compared to name space. Without such analysis, the performance gains risk being perceived as empirical coincidences rather than systematic advantages.

We thank the referee for highlighting the importance of rigorous theoretical justification. Indeed, our proposed statistical arbitrage in rank space strategy has deep theoretical motivations, which justifies (i) enhanced market structure and consequently improved signal-to-noise ratio (Figure 3) and (ii) the robust mean-reversion of residual returns in rank space (Figure 4). Below, we briefly summarize these theoretical foundations.

First, in our ongoing theoretical paper (reference [16] in revised manuscript), we rigorously show why transforming market data into rank space systematically enhances the signal-to-noise ratio. Briefly, consider an indication matrix $\{\theta_t\}_{1 \leq i \leq N, 1 \leq j \leq K} = \chi(c_{i,t} = c_{(k),t})^1$ that maps the stock from name space to rank space based on their ranks in capitalization.

Define the occupation rate $\Theta = \frac{1}{T} \sum_{t=1}^T \theta_t$ which transforms the covariance matrix of stock returns from name space (Σ) into rank space ($\Theta^T \Sigma \Theta$). Under standard ergodic market dynamics assumptions, we theoretically demonstrate that this mapping yields a more structured eigenvalue spectrum, characterized by an enhanced gap between the largest eigenvalue and the smaller eigenvalues. Such an eigenvalue structure inherently enhances the signal-to-noise ratio, directly supporting the empirical findings presented in Figure 3 of our manuscript.

Second, the enhanced mean-reverting behavior of residual returns in rank space (Figure 4) also has deep theoretical anchors. In the appendix B of the revised manuscript, we justify

¹ $c_{i,t}$ and $c_{(k),t}$ are defined in equation (3.5) in the main text. $\chi(\cdot)$ is the indication function that takes 1 if (\cdot) holds otherwise 0.

such behavior under the framework of a hybrid-Atlas model. Starting from the market dynamics in name space where the residual returns in name space are modeled by Brownian motions (equation B.1), theorem B.3 shows that transforming into rank space introduces additional local time terms for the residual returns in rank space, significantly promoting the mean-reversion in the residual returns in rank space.

Unfortunately, theoretical discussions for the first part are more involved which does not fit into the length constraints and primary focus on algorithmic and empirical findings in the current NeurIPS manuscript. In addition, it is a collaborative work with several other researchers which would lead to conflict of interest should we include in the current manuscript. Therefore, we have opted to publish the detailed theoretical derivations separately in a dedicated theoretical manuscript, currently being prepared for submission to a mathematical journal.

The claim that name space suffers from a low signal-to-noise ratio is stated without direct evidence. The authors should provide either empirical diagnostics or theoretical support for this assertion.

We thank the reviewer for raising this important point. The low signal-to-noise ratio in name space is empirically demonstrated in Figure 3 (panels c1-c6), where illustrates that market dynamics in name space are driven by multiple significant factors, leading to a reduced signal-to-noise ratio. To make this connection clearer, we have revised our manuscript (line [189]) by explicitly stating that multiple factor dominance implies a “consequently low signal-to-noise ratio”. In addition, we’ve incorporated references [9], [23]) to highlight the central challenge in financial mathematics of developing robust algorithm within low signal-to-noise ratio environment.

The evaluation assumes a relatively low transaction cost (2 bps). The strategy’s sensitivity to higher costs or slippage, which are common in real-world deployments, is not discussed.

We thank the referee for bringing up this important point. In response, we’ve clarified this sensitivity explicitly in line [222-223], directing readers to Appendix G.3, and specifically

Figure 13(b), where we demonstrate the dependence of Sharpe ratio on the transaction costs. As illustrated, profitability diminishes significantly at transaction costs of around 5 bps.

We acknowledge that the current sensitivity to higher transaction costs primarily arises from the relatively simplistic intra-day rebalancing strategy described in Appendix D in the revised manuscript, which aims at realizing returns in rank space (Equation 3.5).

Developing more sophisticated methods for return realization—such as those based on stochastic control or reinforcement learning—is an important direction for future research, which we believe could significantly enhance strategy performance under higher trading costs.

The paper does not address the computational burden or latency risks introduced by frequent re-ranking of stocks in rank space. This is particularly important for live trading systems.

We thank the referee for highlighting this important aspect. To address this explicitly, we have added the following clarification at line [337-343]:

“Each neural network training iteration takes approximately two hours to converge, resulting in roughly 130 computational hours for back testing portfolio performance from 2007 to 2022, assuming quarterly neural network retraining. In our rank-space statistical arbitrage strategy, the daily portfolio weights in rank space are precomputed by forward propagation of the neural network before market opening, and remain fixed during the trading day. To handle rank changes, the intraday rebalancing of portfolio weights from rank space to name space occurs every 225 minutes.”

Since, once trained, the computational overhead involves primarily updating portfolio weights by ranking stocks according to market capitalization every 3.75 hours, our strategy does not impose significant computational demands or latency risks for real-time deployment.

The experiments are limited to U.S. equities and do not explore whether rank space generalizes to other asset classes (e.g., fixed income) or tasks (e.g., portfolio optimization or risk modeling).

We thank the reviewer for raising this insightful point. Our statistical arbitrage strategy leverages the robust market structure obtained by anonymizing stock-specific details through capitalization-based rank representation, which naturally generalizes to other equity markets.

However, extending rank-space representation to asset classes beyond equities (such as fixed income or derivatives) is less straightforward due to their inherent dimensional complexity (e.g., strike prices and maturities for options, or duration metrics for corporate bonds). Defining a meaningful rank-space representation in these contexts remains an open and interesting direction for future research.

Additionally, while our current work explicitly addresses portfolio optimization (Eq. 3.10), the reviewer's suggestion to explore applications in risk modeling is intriguing and represents a promising avenue for further exploration.

Can the authors provide a theoretical explanation for why rank space leads to a higher signal-to-noise ratio? Is there a geometric or statistical rationale behind this improvement?

We thank the reviewer for bringing up this intriguing question and we believe our previous discussions address the question.

What empirical or theoretical evidence supports the assertion that name space inherently exhibits a lower signal-to-noise ratio? and can it be mitigated?

We thank the reviewer for this insightful question. Empirical evidence supporting the lower signal-to-noise ratio in name space is thoroughly discussed in our earlier response to the second comment on weaknesses, particularly referencing Figure 3 (panels c1–c6).

Mitigating the low signal-to-noise ratio in name space is indeed a central research theme in quantitative finance, with extensive exploration of various techniques, including factor analysis, random matrix theory, and robust statistical learning methods. In this context, our proposed rank-space representation can be viewed as a novel and effective approach that addresses this inherent challenge, aiming specifically to enhance the signal-to-noise ratio by structurally transforming into rank space.

Deep Statistical Arbitrage: Name Space versus Rank Space

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Equity market dynamics are conventionally investigated in name space, where
2 stocks are indexed by company names. However, this perspective often suffers
3 from high volatility and a low signal-to-noise ratio, which poses challenges for
4 effective learning by deep neural networks (DNNs). In contrast, by indexing
5 stocks by their ranks in capitalization, we gain a distinct and more structured
6 view of market behavior in rank space. In this work, we demonstrate that DNNs
7 achieve superior performance in statistical arbitrage when operating in rank space
8 compared to name space. This performance gain is driven by more robust market
9 representations and enhanced mean-reverting properties of residual returns in
10 rank space, which facilitate more efficient learning. Our findings highlight the
11 critical role of domain-informed data transformation in improving deep learning
12 performance in noisy financial environments.

13 **1 Introduction**

14 In equity markets, stocks are conventionally labeled by fixed company identifiers (company names),
15 a perspective that often suffers from high volatility and a low signal-to-noise ratio, hindering effective
16 learning by deep neural networks (DNNs).

17 As an alternative, we consider a rank-space representation, where stocks are indexed by their ranks in
18 capitalization rather than fixed company identifiers. In this view, we focus on the stock at a given
19 rank in capitalization while the corresponding company name may change. This formulation offers a
20 more structured and stable view of market dynamics. We refer to a market labeled by fixed company
21 identifiers (company names) as the *market in name space*, and one labeled by ranks in capitalization
22 as the *market in rank space*.

23 A structured market dynamics is crucial for the performance of statistical arbitrage, a trading strategy
24 that exploits temporary under- and over-pricing of stocks. Specifically, statistical arbitrage constructs
25 market-neutral portfolios that reproduce the residual returns – returns not explained by market factors
26 – and generates profits if these residual returns are mean-reverting.

27 In this paper, we show that operating statistical arbitrage in rank space by DNNs significantly
28 outperforms the name-space counterpart. Our DNN-based rank-space portfolios achieve an average
29 annual return 35.68% and an average Sharpe ratio of 3.28 from 2007 to 2022, accounting for a
30 2-basis-point transaction cost. In contrast, applying the same DNNs in name space yields negligible
31 returns over the same period.

32 Further comparison shows that DNN-based rank-space portfolios not only exploit the importance of
33 mean-reversion during mean-variance optimization, but also implement more intelligent strategies
34 than traditional parametric model by applying flexible leverage and minimizing carry risk. We
35 attribute these improvements to two key advantages of rank space: (i) a more robust and stable

36 representation of market structure, and (ii) enhanced mean-reverting behavior in residual returns.
 37 Together, these properties enable DNNs to extract actionable signals from noisy financial data.
 38 Our findings highlight the importance of domain-informed data representations in financial machine
 39 learning. In particular, they show how appropriate input transformation can significantly improve
 40 the learning efficiency and performance of deep models in complex, noisy environments like equity
 41 markets[9, 23].
 42 The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3
 43 formulates the framework for statistical arbitrage in both name space and rank space. Section 4
 44 presents our empirical results using U.S. equity date. Section 5 concludes with a discussion of
 45 implications and future directions.

46 2 Related work

47 Our results add to the burgeoning literature on machine learning applications in finance. Gu *et al.*
 48 systematically compare various machine learning methods for predicting stock returns[13]. Aceri *et*
 49 *al.* utilize deep reinforcement learning for portfolio managements[1]. Araci explores the correlation
 50 between financial news and stock returns by BERT language model[2]. Horvath *et al.* apply deep
 51 neural networks to calibrate volatility surface in fixed-income markets[17].

52 Our research advances the understanding of statistical arbitrage. While this problem can be framed as a
 53 stochastic control problem solvable via the Hamilton-Jacobi-Bellman (HJB) equation[10, 20, 25, 29],
 54 practical applications are often limited by calibration challenges and the absence of selection mecha-
 55 nism in diverse markets. To address these issues, Avellaneda and Lee[3] propose a pragmatic trading
 56 strategy based on Ornstein–Uhlenbeck (OU) processes, later refined by Yeo and Papanicolaou[28].
 57 More recently, Mulvey *et al.*[22] combine the HJB framework with feed-forward networks, while Kim
 58 *et al.*[19] apply deep reinforcement learning to optimize pair trading strategies. Guijarro-Ordonez *et*
 59 *al.*[4] provide a comprehensive survey of deep learning models for statistical arbitrage.

60 Our paper also enriches the emerging literature on rank-based market models. Fernholz *et al.*, in
 61 their seminal work on stochastic portfolio theory, introduce the concept of functionally generated
 62 portfolios in both name and rank spaces[1]. Building on this foundation, Banner *et al.*[7] and Ichiba
 63 *et al.*[18] study the dynamics of rank capitalizations and gap processes in Atlas and hybrid Atlas
 64 models, respectively. Healy *et al.*[16] systematically investigate the market structure in rank space by
 65 Θ -transformation.

66 3 Formulation

67 This section formulates statistical arbitrage in name and rank space. A schematic overview is shown
 68 in Fig. 1, detailed algorithmic pseudocode is provided in Appendix A.3, and the code is available at
 69 <https://github.com/Infi-Yingfei-Li/stats-arb-rank-space>.

70 3.1 Market decomposition

71 3.1.1 Name space

72 In a market consisting of N stocks, we denote the dividend-adjusted return¹ on stock i at trading day
 73 t by $r_{i,t}$. We adopt a standard factor model for stock returns,

$$r_t - r_f = \beta_t F_t + \epsilon_t, \quad t = 1, 2, \dots, T. \quad (3.1)$$

74 Here, $r_t = \{r_{i,t}\}_{i=1}^N \in \mathbb{R}^N$ are the dividend-adjusted daily return, $r_f \in \mathbb{R}$ is the risk-free rate²,
 75 $F_t \in \mathbb{R}^{K \times 1}$ are the underlying factors, $\beta_t \in \mathbb{R}^{N \times K}$ are the corresponding loadings on K factors,
 76 and $\epsilon_t \in \mathbb{R}^N$ are the residual returns.

77 Without loss of generality, these factors can be written as portfolios of stocks,

$$F_t = \omega_t(r_t - r_f), \quad (3.2)$$

¹The daily return of a stock that accounts for both price changes and dividend payments.

²The investment return with zero-risk financial loss, chosen as the rate of 1-month U.S. treasury bill here.

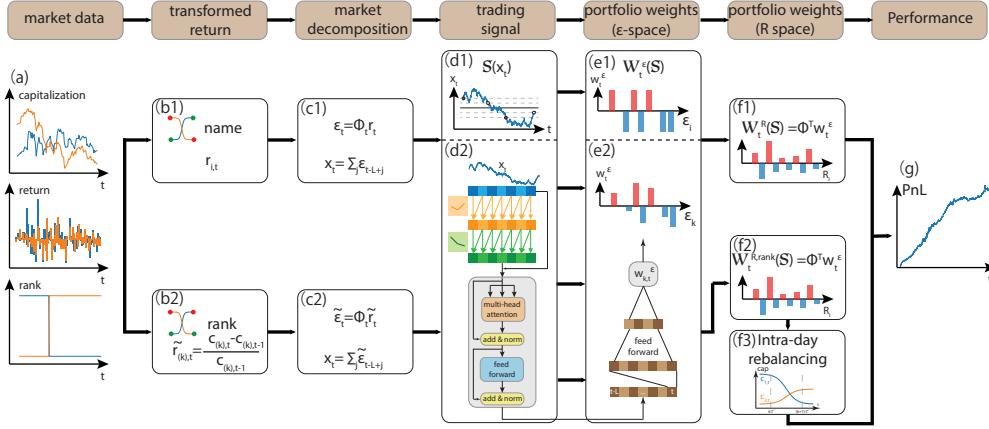


Figure 1: Schematic of the statistical arbitrage algorithm in name space and rank space.

78 where $\omega_t \in \mathbb{R}^{K \times N}$ specifies the factor portfolio weights. Combining (3.1) and (3.2) yields

$$r_t - r_f = \beta_t \omega_t (r_t - r_f) + \epsilon_t \Rightarrow \epsilon_t = (I - \beta_t \omega_t)(r_t - r_f) := \Phi_t(r_t - r_f) \quad (3.3)$$

79 where $\Phi_t := (I - \beta_t \omega_t)$ defines a linear transformation from returns to residual returns. Each $\epsilon_{i,t}$
80 can thus be interpreted as the return of a tradable portfolio with weights given by the i -th row of Φ_t .

81 Consequently, we refer to the space spanned by r_t as the *name equity space*, and the space spanned
82 by ϵ_t as the *name residual space*. We denote the portfolio weights in name equity space as $w_t^{R,\text{name}}$
83 and portfolio weights in name residual space as $w_t^{\epsilon,\text{name}}$. These are related by

$$w_t^{R,\text{name}} = \Phi_t^T w_t^{\epsilon,\text{name}} \quad (3.4)$$

84 Importantly, given any $w_t^{\epsilon,\text{name}}$, the derived $w_t^{R,\text{name}}$ are market neural (proofs in the Appendix C).

85 3.1.2 Rank space

86 We begin by introducing the key variables that characterize the market dynamics in rank space: the
87 daily return on rank k at day t in the continuous-time limit, defined as

$$\tilde{r}_{(k),t} := \frac{c_{(k),t} - c_{(k),t-1}}{c_{(k),t-1}} = \frac{c_{\mathcal{I}_{(k),t},t} - c_{\mathcal{I}_{(k),t-1},t-1}}{c_{\mathcal{I}_{(k),t-1},t-1}}, \quad (3.5)$$

88 where $c_{i,t}$ is the capitalization of stock i at day t , $c_{(k),t}$ is the capitalization of the stock occupying
89 the k -th rank in descending order at day t . $\mathcal{I}_{(k),t}$ represents the company occupying the rank k at day
90 t , and conversely, $\mathcal{R}_{i,t}$ gives the capitalization rank of stock i . Our definition of return in rank space
91 is motivated by the log capitalization process in the hybrid-Atlas model in Appendix B [18].

92 Importantly, \tilde{r}_t does not necessarily correspond to direct financial quantity, as $\mathcal{I}_{(k),t}$ and $\mathcal{I}_{(k),t-1}$ may
93 differ – meaning the stock occupying rank k can change between days. To realize \tilde{r}_t in practice, we
94 develop an intra-day rebalancing strategy, detailed in Appendix D.

95 Following the construction in name space, we assume a factor model for \tilde{r}_t ,

$$\tilde{r}_t - r_f = \tilde{\beta}_t \tilde{F}_t + \tilde{\epsilon}_t, \quad (3.6)$$

96 where $\tilde{r}_t = \{\tilde{r}_{(k),t}\}_{k=1}^N \in \mathbb{R}^N$, $\tilde{\beta}_t \in \mathbb{R}^{N \times K}$, $\tilde{F}_t \in \mathbb{R}^{K \times 1}$, and $\tilde{\epsilon}_t \in \mathbb{R}^N$. Analogously, we model the
97 factors as portfolios of rank returns

$$\tilde{r}_t - r_f = \tilde{\beta}_t \tilde{\omega}_t (\tilde{r}_t - r_f) + \tilde{\epsilon}_t \Rightarrow \tilde{\epsilon}_t = (I - \tilde{\beta}_t \tilde{\omega}_t)(\tilde{r}_t - r_f) := \tilde{\Phi}_t(\tilde{r}_t - r_f), \quad (3.7)$$

98 where $\tilde{\Phi}_t := (I - \tilde{\beta}_t \tilde{\omega}_t)$.

99 We refer to the space spanned by \tilde{r}_t as *rank equity space* and the space spanned by $\tilde{\epsilon}_t$ as *rank residual*
100 space, mirroring our definition in the name space. Let $w_t^{R,\text{rank}}$ and $w_t^{\epsilon,\text{rank}}$ denote the portfolio weights
101 in rank equity space and rank residual space, respectively, related by

$$w_t^{R,\text{rank}} = \tilde{\Phi}_t^T w_t^{\epsilon,\text{rank}}, \quad (3.8)$$

102 As in name space, any $w_t^{\epsilon,\text{rank}}$ generates a market-neutral portfolio $w_t^{R,\text{rank}}$ (proof in Appendix C).

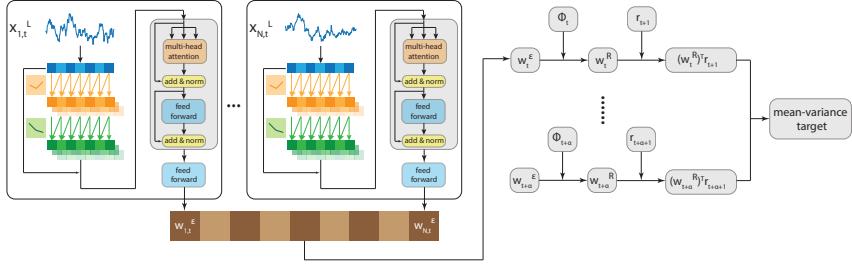


Figure 2: Schematic for the architecture of deep neural networks in both name and rank space.

103 3.2 Trading signals and portfolio weights

104 After computing the residual returns ϵ_t^3 , we derive the cumulative residual returns over a look-back
 105 window of length L :

$$106 x_t^L = (x_{t-L+1}, x_{t-L+2}, \dots, x_t), \quad (3.9)$$

107 where $x_{t-L+\alpha} = \sum_{j=1}^{\alpha} \epsilon_{t-L+j}$, $\alpha = 1, 2, \dots, L$.

108 We adopt DNNs $\mathcal{N} : x_t^L \rightarrow w_t^{\epsilon|\text{NN, name/rank}}$ as a data-driven method to calculate portfolio weights in
 109 residual space w_t^{ϵ} for both name space and rank spaces. The DNNs consist of convolutional layers to
 110 capture local patterns, followed by transformer encoder layers to model global dependencies. The
 neural networks are trained via mean-variance optimization,

$$\begin{aligned} \text{Maximize}_{\mathcal{N}(\cdot)} \quad & \mathbb{E}[(w_t^{R|\text{NN, name/rank}})^T (r_{t+1} - r_f)] - \gamma \text{Var}[(w_t^{R|\text{NN, name}})^T (r_{t+1} - r_f)] \\ \text{s.t.} \quad & w_t^{R|\text{NN, name/rank}} = \frac{\Phi_t^T w_t^{\epsilon|\text{NN, name/rank}}}{\|\Phi_t^T w_t^{\epsilon|\text{NN, name/rank}}\|_1} \\ & w_t^{\epsilon|\text{NN, name/rank}} = \mathcal{N}(x_t^L), \end{aligned} \quad (3.10)$$

111 where γ is the risk-aversion factor. We show a schematic for our DNNs architecture in Fig. 2 with
 112 architecture and implementation details in Appendix A.2.

113 For comparison, we benchmark the DNN performance against a classical parametric model based
 114 on Ornstein-Uhlenbeck (OU) process [3, 28], with the model formulation and execution details in
 115 Appendix E.

116 3.3 Intraday rebalancing

117 The portfolio weights calculated in the rank space are assigned to artificial financial instruments
 118 that yield rank returns in continuous-time limits defined in (3.5). To make the constructed portfolio
 119 practically implementable, it is necessary to convert these portfolio weights into stock-based portfolios
 120 in name space. To address the issue, we propose an intraday rebalancing mechanism.

121 Formally, given the predetermined portfolio weights in rank equity space $\{w_{(k),t}^{\text{rank}}\}_{k=1}^N$ before the
 122 market opening, our goal is to rebalance the portfolio at fixed time intervals of \mathcal{T} minutes such that
 123 the portfolio becomes $\{w_{(k),t}^{\text{rank}} (1 + \tilde{r}_{(k),t+1})\}_{k=1}^N$ by market close, at the sacrifice of additional costs.
 124 To facilitate this discussion, we introduce two processes:

- 125 (i) $w_{(k),t+\tau}^{\text{rank}}$: the dollar-valued portfolio weight for rank k at time $t + \tau$;
- 126 (ii) $w_{i,t+\tau}^{\text{name}}$: the dollar-valued portfolio weight for stock i at time $t + \tau$.

127 Here, t denotes the daily time tick and τ represents the intraday time tick. For instance, for $t = \text{Jan, 3rd, 2022}$ (end of the day) and $\tau = 45$ minutes, $t + \tau$ refers to Jan, 4th, 2022 00:45 AM, and $t + 1$
 128 refers to the end of the day on Jan, 4th, 2022.

³For simplicity, we take a unified notation ϵ_t to denote the residual returns in both name and rank space in the following discussions unless otherwise specified.

130 $w_{(k),t+\tau}^{\text{rank}}$ is the portfolio weight on the k -th rank that evolves strictly based on the rank returns in
 131 continuous-time limit,

$$w_{(k),t+\tau}^{\text{rank}} = w_{(k),t}^{\text{rank}}(1 + \tilde{r}_{(k),t+\tau}), \quad (3.11)$$

132 where $\tilde{r}_{(k),t+\tau} = \frac{c_{(k),t+\tau}}{c_{(k),t}} - 1$. Here, $c_{(k),t+\tau}$ denotes the capitalization at k -th rank at time $t + \tau$.

133 In contrast, $w_{i,t+\tau}^{\text{name}}$ is the portfolio weight on the i -th stock, evolving according to the following rules:

134 (i) Between the rebalancing interval when $t + j\mathcal{T} < t + \tau \leq t + (j+1)\mathcal{T}, j \in \mathbb{N}$,

$$w_{i,t+\tau}^{\text{name}} = w_{i,t+(j\mathcal{T})^+}^{\text{name}} \times \frac{c_{i,t+(j\mathcal{T})^+}}{c_{i,t+(j\mathcal{T})^+}}, \quad (3.12)$$

135 where $(j\mathcal{T})^+ := \lim_{\delta \downarrow 0} (j\mathcal{T} + \delta)$.

136 (ii) At the re-balancing points when $\tau = ((j+1)\mathcal{T})^+, j \in \mathbb{N}$, adjust the portfolio weights via active
 137 trading such that

$$w_{i,t+((j+1)\mathcal{T})^+}^{\text{name}} = \sum_{k=1}^n w_{(k),t+(j+1)\mathcal{T}}^{\text{rank}} \mathbf{1}_{\{\mathcal{R}_{i,t+(j+1)\mathcal{T}}=k\}}. \quad (3.13)$$

138 In other words, we carry out the conversion of portfolio weights between name space and rank space
 139 at the rebalancing points $\tau = ((j+1)\mathcal{T})^+, j \in \mathbb{N}$ through active trading. Notably, the value on the
 140 trading book before trading at $t + (j+1)\mathcal{T}$ is $\sum_{k=1}^N w_{i,t+(j+1)\mathcal{T}}^{\text{name}}$, while the desired value immediately
 141 after trading is $\sum_{i=1}^N w_{i,t+(j+1)\mathcal{T}}^{\text{name}} = \sum_{k=1}^N w_{(k),t+(j+1)\mathcal{T}}^{\text{rank}}$. The two values are not necessarily equal
 142 when there exists switching of ranks in capitalization between $t + j\mathcal{T}$ and $t + (j+1)\mathcal{T}$ (elaborated
 143 by a case study in Appendix D and by the hybrid-Atlas model in Appendix B). Consequently, the
 144 cost of the active trading at $t + (j+1)\mathcal{T}$ involves two components,

$$\begin{aligned} \text{cost}(t + (j+1)\mathcal{T}^+; w_{(k),t}^{\text{rank}}) = & \left(\sum_{i=1}^N w_{i,t+(j+1)\mathcal{T}^+}^{\text{name}} - \sum_{k=1}^N w_{(k),t+(j+1)\mathcal{T}}^{\text{rank}} \right) \\ & + \eta \sum_{i=1}^N |w_{i,t+(j+1)\mathcal{T}^+}^{\text{name}} - w_{i,t+(j+1)\mathcal{T}}^{\text{name}}|, \end{aligned} \quad (3.14)$$

145 where η is the transaction cost factor. Here, the cost at time $t + (j+1)\mathcal{T}^+$ depends on the portfolio
 146 weights assigned at the beginning of the trading day, $w_{(k),t}^{\text{rank}}$, because the intraday portfolio weights
 147 $w_{i,t+\tau}^{\text{name}}$ and $w_{(k),t+\tau}^{\text{rank}}$ are recursively governed by the system dynamics (3.11), (3.12), and (3.13))
 148 from the initial condition $w_{(k),t}^{\text{rank}}$. We highlight this dependence by including the $w_{(k),t}^{\text{rank}}$ as a parameter
 149 for the cost in (3.14). We refer to the first term in (3.14) as latency cost, the second term as cost
 150 from the bid-ask spread, with their sum representing the total transaction cost. The terminology are
 151 rationalized in the case study in Appendix D.

152 The precise implementation of the intraday rebalancing is summarized in Appendix D Algorithm
 153 4 along with a schematic in panel (f3) in Fig. 1. For our backtesting, we primarily use $\eta = 2$ basis
 154 points to account for the cost from the bid-ask spread. This setting approximately corresponds to a
 155 5-10 cents bid-ask spread for our investment universe, the top 500 stocks in the U.S. equity market.

156 3.4 Backtesting

157 We evaluate the portfolio performance by calculating the historical profit-and-loss (PnL) V_t and the
 158 Sharpe ratio.

159 For portfolios in name space,

$$V_{t+1} = (1 + r_{f,t+1}) \times (V_t - \sum_i \Lambda w_{i,t} V_t - \text{TC}) + \sum_i \Lambda V_t w_{i,t} (1 + r_{i,t+1}), \quad (3.15)$$

160 where $\Lambda = 1$ is the leverage, $r_{f,t+1}$ is the risk-free rate during the trading day $t + 1$, and w_t are
 161 normalized by l_1 norm. The transaction cost is given by $\text{TC} = \eta \sum_i \Lambda |V_t w_{i,t} - V_{t-1} w_{i,t-1} (1 + r_{i,t})|$,
 162 where η is the transaction cost factor, set to 2 basis points.

163 For portfolios in rank space, the PnL evolves as

$$V_{t+1} = (V_t - \sum_{k=1}^N w_{(k),t}^{R,\text{rank}})(1 + r_{f,t+1}) + \sum_{k=1}^N \Lambda V_t w_{(k),t}^{R,\text{rank}}(1 + \tilde{r}_{(k),t+1}) - \sum_{j:t < t+j\mathcal{T} \leq t+1} \text{cost}(t + j\mathcal{T}^+; \Lambda V_t w_{(k),t}^{R,\text{rank}}), \quad (3.16)$$

164 where the last term accounts for the transaction costs due to intraday rebalancing defined in
 165 (3.14), where we substitute generic portfolio weights $w_{(k),t}^{\text{rank}}$ in (3.14) into specific portfolio weights
 166 $\Lambda V_t w_{(k),t}^{R,\text{rank}}$ in (3.16).

167 4 Empirical results for the U.S. equities

168 4.1 Data and experimental setup

169 We collect dividend-adjusted daily return, price, shares outstanding, and capitalizations for the U.S.
 170 securities from Center for Research in Security Prices (CRSP), covering January 1990 to December
 171 2022. Intraday price data at 1-minute resolution from January 2005 to December 2022 are obtained
 172 from *Polygon.io*. We construct intraday capitalization data by combining CRSP shares outstanding
 173 with *Polygon.io* intraday prices. The one-month Treasury bill rate from the Kenneth French Data
 174 Library is used as the risk-free rate r_f . Further implementation details are provided in Appendix A.

175 4.2 Market structure: name space versus rank space

176 We begin by comparing the market dynamics in name space and rank space, highlighting two key
 177 advantages of rank space: (i) a more structured market dynamics in rank space, and (ii) a more
 178 enhanced mean-reverting behavior of residual returns in rank space – both critical motivations for
 179 operating statistical arbitrage in rank space.

180 First, we show the market capitalization across ranks, averaged over five-year window from 1991
 181 to 2022 in Fig. 3(a), revealing a stable distribution in rank space [11]. A principal component
 182 analysis (PCA) on the correlation matrix suggests significantly larger leading eigenvalue in rank space
 183 compared to that in name space (Fig. 3(b)), implying that a greater proportion of market variance is
 184 captured by the first eigenvector in rank space. This points to a more structured and concentrated
 185 market dynamic in rank space.

186 More importantly, the single-factor structure in rank space substantially simplifies market decom-
 187 position. We present the empirical eigenvalue spectra of the correlation matrix in both spaces
 188 across different periods in Fig. 3(c1-c6) and (d1-d6). In name space, several eigenvalues exceed
 189 the Marchenko–Pastur upper bound [4], indicating a multi-factor market and consequently low
 190 signal-to-noise ratio (Fig. 3(c1-c6)). In contrast, rank space exhibits a sharp bulk-edge separation with
 191 a dominant single factor (Fig. 3(d1-d6)). This clearer structure enables more well-defined separation
 192 of market factors from residuals during market decomposition.

193 Second, we observe markedly faster mean-reversion of residual returns in rank space, critical for
 194 statistical arbitrage. We quantify mean-reversion by fitting cumulative residual returns x_t^L to an OU
 195 process and extracting the mean-reversion time τ . Fig. 4 shows the empirical distribution of τ over
 196 five-year windows from 1991 to 2022. In name space (Fig. 4(a1–a6)), the distribution is heavy-tailed
 197 toward large τ , reflecting slower mean reversion. In contrast, rank space (Fig. 4(b1–b6)) exhibits a
 198 sharper concentration in the fast mean-reverting regime, with fewer instances of slow mean-reversion
 199 ($\tau > 30$ days, shaded area). A non-parametric analysis in Appendix F yields consistent results.

200 4.3 Portfolio performance

201 We present the PnL V_t in Fig. 5 with portfolio weights w_t^R are calculated by four scenarios: (i) the
 202 parametric benchmark model in name space in panel (a), (ii) the parametric benchmark model in
 203 rank space in panels (b,c), (iii) DNNs in name space in panel (d), and (iv) DNNs in rank space in
 204 panels (e, f). The corresponding Sharpe ratios from 2016 to 2022 are summarized in Fig. 5(g, h), with
 205 year-by-year statistics provided in Appendix G.1 (without transaction costs: Table 1; with transaction
 206 costs: Table 2).

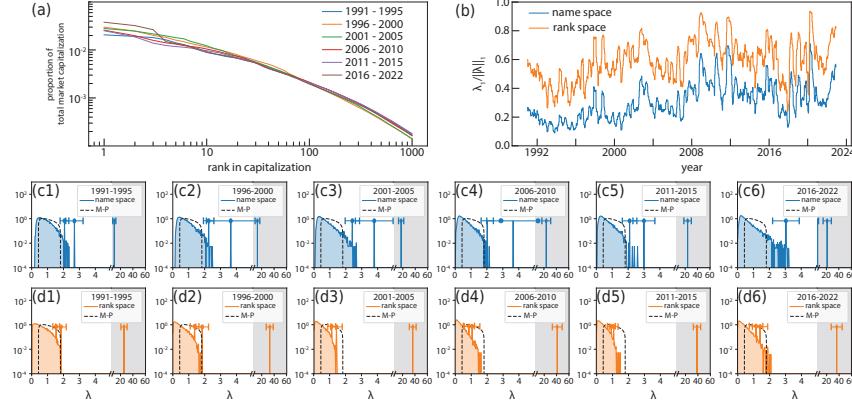


Figure 3: **Market structure in name space versus rank space.** (a) Proportion of total market capitalization versus ranks in capitalization. (b) The principal eigenvalue of the correlation matrices of r_t in name space (blue) and for \tilde{r}_t rank space (orange). (c, d) The empirical probability distribution density of the eigenvalue spectrum of the correlation matrices versus Marchenko-Pastur distribution. The market exhibits a more structured correlation in rank space compared to name space.

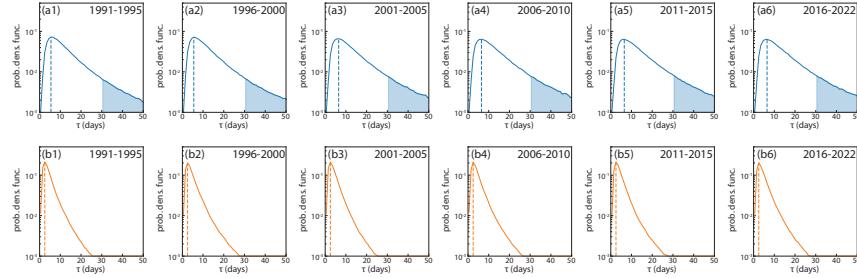


Figure 4: **Mean-reverting time τ in name space versus rank space.** The τ is evaluated by fitting the cumulative residual return x_t^L to an Ornstein–Uhlenbeck(OU) process. (a1-a6) The empirical distributions of mean-reverting time τ in name space, with maximum empirical probability at ~ 6 days (vertical dashed lines). (b1-b6) The empirical distributions of τ in rank space, with maximum empirical probability at ~ 2.5 days (vertical dashed lines). The residual returns in rank space show faster mean-reverting behavior, favorable for statistical arbitrage.

207 The traditional statistical arbitrage strategy in name space using the parametric model exhibits
 208 diminishing profitability after the 2010s. In rank space, the parametric model yields mixed results:
 209 while initial performance without transaction costs appears attractive (Fig. 5(b)), accounting for
 210 transaction costs ($\eta = 0.0002$) leads to a monotonic decline in PnL (Fig. 5(c)). This stark contrast
 211 highlights the substantial cost of realizing continuous-time rank returns \tilde{r}_t via intraday rebalancing.

212 To address these challenges, we leverage DNNs to better exploit market patterns, particularly in
 213 rank space. The performance of DNNs in name and rank spaces diverges sharply. In name space,
 214 DNNs fail to improve returns or Sharpe ratios (Fig. 5(a, d); Appendix G.1, Table 1, Table 2). In
 215 contrast, in rank space, DNNs substantially enhance portfolio performance, achieving an average
 216 annual return of 35.68% and an average Sharpe ratio of 3.28 from 2007 to 2022 (Fig. 5(h); Table 2),
 217 even after accounting for transaction costs. This success is driven by the effective exploitation of
 218 mean-reversion behavior in rank space by DNNs that yields an average annual return of 206.49% and
 219 an average annual Sharpe ratio of 9.04 without transaction costs (Appendix G.1, Fig. 5(g), Table 1),
 220 sufficient to offset the substantial costs in intraday rebalancing to realize \tilde{r}_t .

221 Further characterization of the portfolios is provided in Appendix G. Appendix G.1 presents annu-
 222 alized performance statistics, Appendix G.2 examines market and dollar neutrality, Appendix G.3
 223 discusses dependence on transaction costs, and Appendix G.4 analyzes the role of rank-swapping
 224 timescales on portfolio performance.

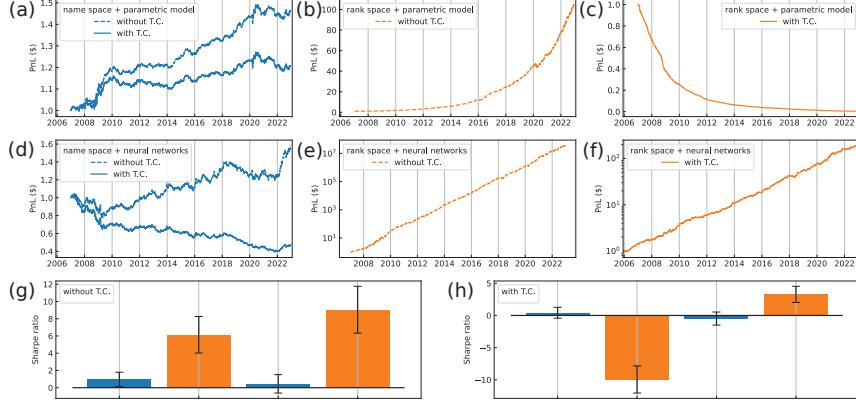


Figure 5: Summary of portfolio performance. The PnL dynamics V_t are computed using (3.15) in name space and (3.16) in rank space. **(a)** PnL using portfolio weights from the parametric model in name space; dashed and solid lines represent results without and with transaction costs, respectively. **(b, c)** PnL using portfolio weights from the parametric model in rank space; dashed/solid lines represent results without/with transaction costs in panel (b)/(c). **(d)** Same as (a), but using portfolio weights derived from DNNs in name space. **(e, f)** Same as (b, c), but using portfolio weights from DNNs in rank space. **(g, h)** Average Sharpe ratio without/with transaction costs shown in panel (g)/(h). Notably, portfolios derived from DNNs in rank space perform significantly better than those in name space.

225 4.4 The intelligence inside the neural networks

226 To understand the outperformance of DNNs compared to the benchmark parametric model in rank
227 space, we analyze the relationship between the input (the trajectories of the cumulative residual return
228 x_t^L) and the output (the portfolio weights in residual space w_t^e).

229 We parameterize x_t^L by two key variables: the deviation from long-term average $\frac{x_t - \mu}{\sigma}$, and the
230 mean-reverting time τ , following the trading signal suggested by the parametric model (E.4). Each
231 x_t^L thus corresponds to a point in the plane spanned by $\frac{x_t - \mu}{\sigma}$ and τ , color-coded by w_t^e (Fig. 6).

232 We evaluate four scenarios: (i) parametric model in name space (Fig. 6(a)), (ii) DNNs in name space
233 (Fig. 6(b)), (iii) parametric model in rank space (Fig. 6(d)), and (iv) DNNs in rank space (Fig. 6(e)).
234 We also report the average holding periods before liquidation for each method (Fig. 6(c, e)).

235 Compared to the parametric benchmark, the DNNs demonstrate more sophisticated trading behavior
236 in rank space. Despite operating directly on raw cumulative return trajectories, the DNNs successfully
237 uncover the significance of mean-reversion through mean-variance optimization (Fig. 6(e)). In
238 addition, the DNNs improve the execution strategy along three dimensions:

239 (i) Variable leverage on deviations. The DNNs assign higher leverage to positions with larger
240 normalized deviations $\frac{x_t - \mu}{\sigma}$, enhancing profit potential during significant market moves (Fig. 6(e)).

241 (ii) Flexible opportunity thresholds. Rather than relying on rigid mean-reversion time (τ) cutoffs as
242 in the parametric model (Fig. 6(a, d)), DNNs embrace a broader range of trading opportunities while
243 concentrating investments in trajectories with fast mean-reversion (τ small) (Fig. 6(b, e)).

244 (iii) Shorter holding periods. DNNs reduce average holding periods to around 5 days, compared to
245 approximately 10 days under the parametric model (Fig. 6(c, e)). This minimizes carry-over risk,
246 which is particularly important when employing variable leverage across positions.

247 We also emphasize the critical role of market data preprocessing. While the DNNs achieve substantial
248 improvements in rank space, they fail to converge or deliver gains in name space (Fig. 7; Fig. 5(d)).
249 Despite both input spaces being derived from similar capitalization data, strategic reorganization into
250 rank space dramatically improves training efficiency (Fig. 7) and portfolio performance (Fig. 5). This
251 illustrates the importance of domain-informed data transformations in financial machine learning:

appropriate restructuring of the input space can substantially enhance the learning efficiency and performance of deep models in complex, noisy environments like equity markets.

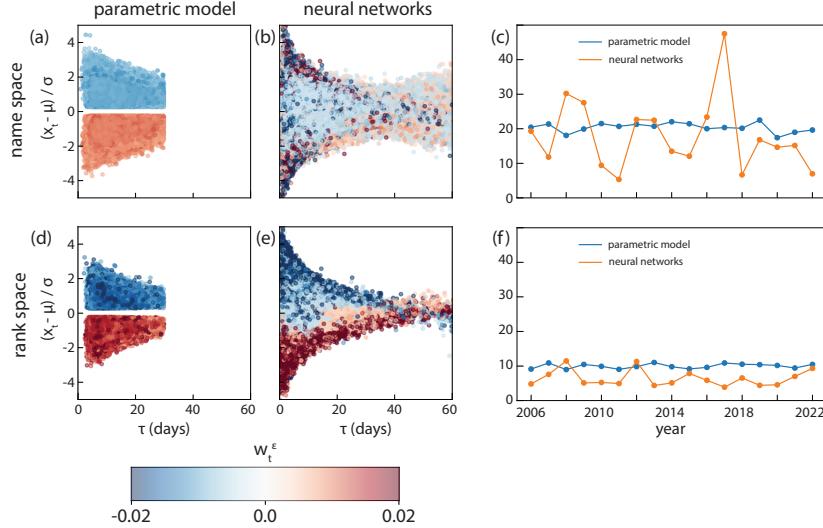


Figure 6: Portfolio weights in residual space: parametric model versus neural networks. (a, b, d, e) We illustrate the behavior of both the parametric model and neural networks by analyzing the relationship between the input, the trajectories of cumulative residual returns x_t^L , and the output, the portfolio weights in residual space, w_t^ϵ . The input x_t^L are parameterized by two variables: (i) deviation from long-term average $\frac{x_t - \mu}{\sigma}$, and (ii) mean-reverting time τ . Each x_t^L thus corresponds to a point in the plane spanned by $\frac{x_t - \mu}{\sigma}$ and τ , color-coded by w_t^ϵ . This analysis is performed with w_t^ϵ calculated by four scenarios: (a) the parametric model in name space; (b) neural networks in name space; (d) the parametric model in rank space; (e) neural networks in rank space. (c, f) Panel (c) and (f) show the average holding time for portfolios derived in name space and rank space, respectively. The DNN-derived portfolios exhibit more intelligence in terms of variable leverage, flexible opportunity thresholds, and shorter holding periods.

5 Limitations and discussion

Our current strategy is sensitive to transaction costs, particularly due to the frequent intraday rebalancing required to realize continuous-time rank returns. While our proposed intraday rebalancing mechanism (Appendix D) provides a baseline, it offers substantial room for further optimization. Future work could formulate this challenge as a stochastic control problem, leveraging physics-informed neural networks [24] to solve the resulting high-dimensional partial differential equations, or alternatively as a reinforcement learning problem based on real-market data.

Beyond financial markets, the rank-based representation introduced here may generalize to other many-particle systems, such as those encountered in many-particle physics [21], biology [8], and social sciences [12].

6 Conclusion

We have introduced a novel statistical arbitrage method that leverages the robust market structure in rank space. Although rank space and name space contain the same underlying information, the significant performance improvements achieved by DNNs in rank space highlight the critical importance of domain-informed data representations in financial machine learning. In particular, our results demonstrate how appropriate transformations of the input space can dramatically enhance the learning efficiency and performance of deep models in complex, noisy environments such as many-particle systems like equity markets.

272 **A Implementation details**

273 **A.1 Calibrating investment universe and market decomposition**

274 Our backtesting spans January 2006 to December 2022, covering both the subprime mortgage crisis
 275 and the highly competitive post-2010 market period. On each trading day after market close, we re-
 276 calibrate the investment universe by selecting stocks that (i) rank among the top 500 in capitalization
 277 as of day t , ensuring sufficient liquidity, and (ii) have valid historical return data available for day
 278 $t + 1$. This selection procedure minimizes potential selection bias to the best of our ability.

279 We then perform principal component analysis (PCA) on the selected returns using a 252-day
 280 lookback window to extract leading eigenvectors as market factors F_t . Specifically, we retain the top
 281 five eigenvectors (associated with the five largest eigenvalues) in name space, and the top eigenvector
 282 in rank space. Factor loadings β_t are estimated using a 60-day lookback window, from which the
 283 transformations Φ_t and residual returns ϵ_t are computed. Cumulative residual returns x_t^L are similarly
 284 evaluated over the same 60-day window and are used as inputs to either the parametric benchmark
 285 model (Appendix E) or the deep neural networks to generate portfolio weights and calculate the
 286 resulting PnL.

287 **A.2 Deep neural networks**

288 We delve into the specific architecture of our neural networks, illustrated in Fig. 2. Our CNN-
 289 transformer architecture harnesses the strengths of CNN in extracting local patterns and transformers
 290 in capturing long-term dependencies. The inputs of our neural networks are the trajectories of
 291 cumulative residual returns, $x_t^L \in \mathbb{R}^{N \times L}$, processed through two layer of multi-channel convolutional
 292 networks, followed by a standard transformer encoder layer that models global relationships via
 293 multi-head attention. Specifically, in the convolutionary layer,

$$\begin{aligned} x_t^{(1)} &= \frac{x_t^L - \mathbb{E}(x_t^L)}{\sqrt{\text{Var}(x_t^L) + \epsilon}} \times \gamma^{(1)} + \beta^{(1)}, & y_t^{(1)} &= W^{(1)} * x_t^{(1)} + b^{(1)}, & z_t^{(1)} &= \text{ReLU}(y_t^{(1)}) + x_t^{(1)}; \\ x_t^{(2)} &= \frac{z_t^{(1)} - \mathbb{E}(z_t^{(1)})}{\sqrt{\text{Var}(z_t^{(1)}) + \epsilon}} \times \gamma^{(2)} + \beta^{(2)}, & y_t^{(2)} &= W^{(2)} * x_t^{(2)} + b^{(2)}, & z_t^{(2)} &= \text{ReLU}(y_t^{(2)}) + x_t^{(2)}. \end{aligned} \quad (\text{A.1})$$

294 The superscript (1) or (2) specifies the layer number. $x_t^{(1)} \in \mathbb{R}^{N \times L}$ is the input of the first convolu-
 295 tional layer. $W^{(1)} \in \mathbb{R}^{D_{\text{channel}} \times 1 \times D_{\text{kernel}}}$ and $W^{(2)} \in \mathbb{R}^{D_{\text{channel}} \times D_{\text{channel}} \times D_{\text{kernel}}}$ are the convolutionary ker-
 296 nels for the convolution operator denoted by $*$, $b^{(1,2)} \in \mathbb{R}^{D_{\text{channel}}}$ is the bias, and $y_t^{(1,2)} \in \mathbb{R}^{N \times D_{\text{channel}} \times L}$
 297 is the output of convolutionary operator. D_{channel} is the number of channels and D_{kernel} is the size of
 298 the convolution kernel. We adopt a rectified linear unit (denoted as $\text{ReLU}(\cdot)$) as our activation function.
 299 We also apply (i) instance normalization [26] with learnable parameter $\gamma^{(1,2)}$ and $\beta^{(1,2)}$ at the input
 300 of each convolution layer to accelerate the training process, and (ii) residual connection [15] to avoid
 301 vanishing gradients by directly connecting the input $x_t^{(1,2)}$ to the output $z_t^{(1,2)} \in \mathbb{R}^{N \times D_{\text{channel}} \times T}$. We
 302 choose the hyper-parameters for our neural networks as number of channels $D_{\text{channel}} = 8$ and size of
 303 the convolution kernel $D_{\text{kernel}} = 2$.

304 The outputs of convolutionary layers, $z_t^{(2)} \in \mathbb{R}^{N \times D_{\text{channel}} \times T}$ are subsequently fed into a standard
 305 transformer encoder layer [27]. The transformer encoder layer utilizes the multi-head attention

306 modeled by the inner product between the famous key-query-value matrices. To elaborate,

$$\begin{aligned}
x_t^{\text{transformer}} &= (z_t^{(2)})^T \\
\begin{cases} Q_i \\ K_i \\ V_i \end{cases} &= \text{DropOut}(W_i^Q x_t^{\text{transformer}} + b_i^Q) \\
&= \text{DropOut}(W_i^K x_t^{\text{transformer}} + b_i^K), \quad i = 1, 2, \dots, H \\
&= \text{DropOut}(W_i^V x_t^{\text{transformer}} + b_i^V) \\
\text{head}_i &= \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_{\text{channel}}/H}}\right), \quad i = 1, 2, \dots, H \\
y_t &= \text{Concat}(\text{head}_1 V_1, \dots, \text{head}_H V_H) \\
z_t &= \text{LayerNorm}(x_t^{\text{transformer}} + y_t) \\
o_t &= \text{LayerNorm}(W^O z_t + b^O + y_t),
\end{aligned} \tag{A.2}$$

307 where $x_t^{\text{transformer}} \in \mathbb{R}^{N \times T \times D_{\text{channel}}}$ is the input of the transformer encoder layer. In addition,
308 $W_i^Q \in \mathbb{R}^{(D_{\text{channel}}/H) \times D_{\text{channel}}}$, $W_i^K \in \mathbb{R}^{(D_{\text{channel}}/H) \times D_{\text{channel}}}$, and $W_i^V \in \mathbb{R}^{(D_{\text{channel}}/H) \times D_{\text{channel}}}$ are the
309 linear weights. $b_i^Q \in \mathbb{R}^{D_{\text{channel}}/H}$, $b_i^K \in \mathbb{R}^{D_{\text{channel}}/H}$, and $b_i^V \in \mathbb{R}^{D_{\text{channel}}/H}$ are the bias. Softmax(\cdot)
310 stands for softmax function and $\text{Concat}(\cdot)$ for matrix concatenation, and $y_t \in \mathbb{R}^{N \times T \times D_{\text{channel}}}$ is the
311 output of multi-attention layer. $W^O \in \mathbb{R}^{D_{\text{channel}} \times D_{\text{channel}}}$ and $b^O \in \mathbb{R}^{D_{\text{channel}}}$ are the linear weights and
312 bias in the output linear layer. $o_t \in \mathbb{R}^{N \times T \times D_{\text{channel}}}$ is the output of the transformer. In addition to the
313 residual connection similar to the convolutional layer, we also introduce the drop-out technique, de-
314 noted as $\text{Dropout}(\cdot)$, to regularize overfitting with drop-out probability p , and layer normalization [5],
315 denoted as $\text{LayerNorm}(\cdot)$, to improve training stability. We choose the hyper-parameters for our
316 neural networks as the number of heads $H = 4$ and the drop-out probability $p = 0.25$.

317 Finally, we choose the last slice along the time axis in the output of the transformer, $o_t \in \mathbb{R}^{N \times T \times D_{\text{channel}}}$
318 as the hidden state summarizing the information up to time t . The portfolio weights
319 in residual space $w_t^{\epsilon|\text{NN}, \text{name/rank}}$ are calculated by a linear relationship,

$$w_t^{\epsilon|\text{NN}, \text{name/rank}} = W^F(o_t[:, -1, :]) + b^F, \tag{A.3}$$

320 where $o_t[:, -1, :] \in \mathbb{R}^{N \times D_{\text{channel}}}$ means the last slice along the second-dimension (time-axis) of o_t , and
321 $W^F \in \mathbb{R}^{1 \times D_{\text{channel}}}$, $b^F \in \mathbb{R}$ are the parameters.

322 For the mean-variance optimization target in (3.10), the empirical expectation and variance are
323 obtained over a consecutive time window of length T ,

$$\begin{aligned}
\mathbb{E}[(w_t^{R|\text{NN}})^T (r_{t+1} - r_f)] &\approx \frac{1}{T} \sum_{\alpha=1}^T (w_{t+\alpha}^{R|\text{NN}})^T (r_{t+\alpha+1} - r_f) \\
\text{Var}[(w_t^{R|\text{NN}})^T (r_{t+1} - r_f)] &\approx \frac{1}{T} \sum_{\alpha=1}^T [(w_{t+\alpha}^{R|\text{NN}})^T (r_{t+\alpha+1} - r_f) - \mathbb{E}((w_t^{R|\text{NN}})^T (r_{t+1} - r_f))]^2
\end{aligned} \tag{A.4}$$

324 We choose the risk-aversion factor $\gamma = 2$ and length of time window $T = 24$ days.

325 The neural networks are trained in two steps. The first step aims at optimizing the hyper-parameters
326 of neural network. Specifically, to evaluate the portfolio weights from the trading day t to day $t + 252$,
327 we utilize the data from $t - 1000$ to $t - 60$ as the training data set and from $t - 59$ to $t - 1$ as the
328 validation data set, from which we determine hyper-parameters from the converged mean-variance
329 target in the training curve (Fig. 7). The hyperparameters of our optimized neural network architecture
330 are elaborated in Appendix A.2.

331 The second step aims at increasing the updating frequency of parameters in the neural networks from
332 annually to quarterly while keeping hyper-parameters fixed. More explicitly, to evaluate the portfolio
333 weights from the trading day t to day $t + 63$, we use the data from trading day $t - 500$ to day $t - 1$ as the
334 training data set. Empirically, increasing the updating frequency is significantly beneficial
335 to the performance of neural network, due to the non-stationarity in financial data. The training
336 tasks utilize PyTorch 2.2.0, and are parallelized on a workstation with a CPU from AMD Ryzen
337 Threadripper Pro 5955 WX and two GPUs from Nvidia GeForce RTX 4090. Each neural network
338 training iteration takes approximately two hours to converge, resulting in roughly 130 computational

339 hours for backtesting portfolio performance from 2007 to 2022, assuming quarterly neural network
 340 retraining. In our rank-space statistical arbitrage strategy, the daily portfolio weights in rank space
 341 are precomputed by forward propagation of the neural network before market opening, and remain
 342 fixed during the trading day. To handle rank changes, the intraday rebalancing of portfolio weights
 343 from rank space to name space occurs every 225 minutes.

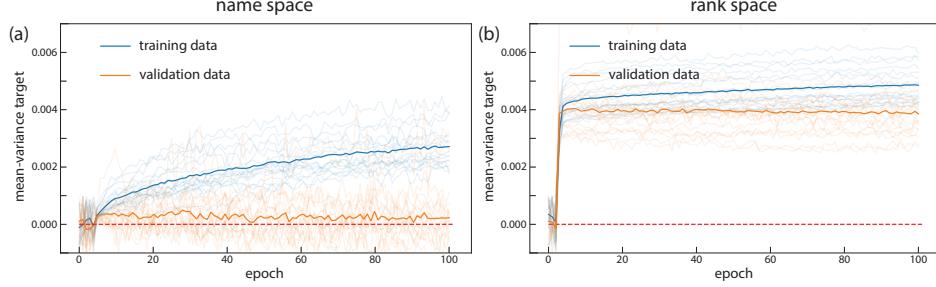


Figure 7: Training curves of neural networks. **(a, b)** We show the mean-variance target as a function of training epochs for neural networks as specified by (3.10) in name space (a) and in rank space (b). The training curves originate from the phase I training process focusing on hyperparameter tuning. To evaluate the out-of-sample portfolio weights from trading day t to $t + 252$, we use the training data from day $t - 1000$ to day $t - 60$ and validation data from day $t - 59$ to $t - 1$ as the validation data. The neural networks are re-trained annually with random initialization, yielding approximately 17 training curves from 2006 to 2022. The transparent lines show individual training curves with their average represented by the opaque lines. The neural networks in rank space are more efficiently trained than those in name space.

345 We provide implementation details in the form of pseudocode. Algorithm 1 performs market
 346 decomposition in both name and rank space, as discussed in section 3.1. Algorithm 2 calculates
 347 the portfolio weights using the parametric model (section E), while algorithm 3 does so via neural
 348 networks (section 3.2 and A.2). Algorithm 4 handles the conversion of portfolio weights between
 349 name space and rank space (section D). Algorithm 5 computes the summary statistics of portfolio
 350 performance reported in 1 and 2. Finally, algorithm 6 and 7 integrate the proposed algorithmic
 351 components above to implement the complete statistical arbitrage strategy in name and rank space,
 352 respectively.

Algorithm 1 Market decomposition (PCA) [Fig. 1 panel(c1, c2)]

Input: $r_t, r_{f,t}, K$
Output: ϵ_t, Φ_t

Function market_decomposition($r_t, r_{f,t}, K$):

Principal component analysis: $r_t - r_{f,t} = U\Sigma V^T$
 $F_t \leftarrow (v_1, v_2, \dots, v_K)$, where v_k is the k -th column of V^T
 Calculate ω_t by solving $F_t = \omega_t(r_t - r_f)$
 Calculate β_t as the coefficient of the linear regression $r_t - r_f \sim F_t$
 $\Phi_t \leftarrow I - \beta_t \omega_t$
 $\epsilon_t \leftarrow \Phi_t(r_t - r_{f,t})$
return ϵ_t, Φ_t

// Input:
 // r_t : return in name space or transformed return in rank space.
 // $r_{f,t}$: risk-free rate at the end of trading day t .
 // K : number of market factors, predetermined by analyzing eigenvalue
 // spectrum of the correlation matrix.
// Output:
 // ϵ_t : residual returns in name space or rank space.
 // Φ_t : transformation between residual space and equity space
 // ((3.1) for name space and (3.6) for rank space).
// Note:
 // The algorithm realizes the formulation in section 2.1.
 // Factors F_t and ω_t are calculated on a 252-day look-back window.
 // Loadings β_t are calculated on a 60-day look-back window.
 // F_t , ω_t , and β_t are updated daily.
 // $K = 5$ for name space and $K = 1$ for rank space based on empirical
 // eigenvalue spectrum of the correlation matrix (Fig. 3(c,d))).

Algorithm 2 Portfolio weights by parametric model [Fig. 1, panel(d1, e1)]

Input: ϵ_t, Φ_t
Output: $w_t^{R|OU}$

Function portfolio_weights_by_parametric_model(ϵ_t, Φ_t)
 Calculate x_t^L by (3.9)
 Estimate $\tau, \mu, \sigma, x_t, R^2$ by fitting x_t^L to an OU process ((E.1))
 Calculate $w_t^{\epsilon|OU}$ by (E.4)
 $w_t^{R|OU} \leftarrow \Phi_t^T w_t^{\epsilon|OU}$ ((E.7))
return $w_t^{R|OU} / \|w_t^{R|OU}\|_1$

// Input:
// ϵ_t : residual returns calculated from Algorithm 1
// Φ_t : transformation matrix between equity space and residual space
from
// Algorithm 1
// Output:
// $w_t^{R|OU}$: l_1 -normalized portfolio weights by parametric model.
// For name space, it stands for the portfolio weights on stocks.
// For rank space, it corresponds to the portfolio weights on
// artificial financial instruments that realize rank returns
// defined in (3.5).
// Note:
// The algorithm realizes the formulation in section 2.2.1.
// τ, μ, σ are fitting parameters of OU process.
// Risk control by ignoring $\tau > 30$ days (E.4).

Algorithm 3 Portfolio weights by neural networks [Fig. 1, panel(d2, e2)]

Input: ϵ_t, Φ_t
Output: $w_t^{R|NN}$

Function portfolio_weights_by_neural_networks(ϵ_t, Φ_t)
 Calculate x_t^L by (3.9)
 Train neural network in-sample for mean-variance optimization ((3.10))
 Calculate $w_t^{\epsilon|NN}$ out-of-sample from trained neural network
return $w_t^{R|NN} / \|w_t^{R|NN}\|_1$

// Input:
// ϵ_t : residual returns calculated from Algorithm 1.
// Φ_t : transformation matrix between equity space and residual space
from
// Algorithm 1.
// Output:
// $w_t^{R|NN}$: l_1 -normalized portfolio weights by parametric model.
// For name space, it stands for the portfolio weights on stocks.
// For rank space, it corresponds to the portfolio weights on
// artificial financial instruments that realize rank returns.
// defined in (3.5).
// Note:
// The algorithm realizes the formulation in section 2.2.2.
// No pre-screening on trading opportunities x_t^L applied.
// Neural networks internally prioritize various trading opportunities
// and manage risk (Fig. 6).
// The mean-variance target is evaluated on a 24-day window.

Algorithm 4 intraday rebalancing [Fig. 1 panel (f3)]

Input: $w_t^R, r_{f,t}, \mathcal{T}$
 $c_{t+\tau}, \quad t = 1, 2, \dots, T$ (days), $\tau = 1, 2, \dots, N$ (minutes)

Output: $V_t, \quad t = 1, 2, \dots, T + 1$

Function intraday_rebalancing($w_t^R, r_{f,t}, \mathcal{T}, c_{t+\tau}$)

$t \leftarrow 0, V_t \leftarrow 1, w^{\text{prev}} \leftarrow 0$

$\triangleright T$ is daily time tick

While $t \leq T$

$w_{(k),t}^{\text{rank}} \leftarrow w_{(k),t}^R, \quad k = 1, 2, \dots, N$

$w_{\mathcal{I}_{(k),t},t}^{\text{name}} \leftarrow w_{(k),t}^R, \quad k = 1, 2, \dots, N$ $\triangleright \mathcal{I}_{(k),t}$ maps from rank to name

$V_t \leftarrow V_t - \sum_i w_{i,t}^{\text{name}} - 0.0002 \times \|w_t^{\text{name}} - w^{\text{prev}}\|_1$

$\tau \leftarrow 0$

While $t + \tau < (t + 1)$ $\triangleright \tau$ is intraday time tick

$w_{(k),t+\tau}^{\text{rank}} \leftarrow w_{(k),t+\tau-1}^{\text{rank}} \times \frac{c_{(k),t+\tau}}{c_{(k),t+\tau-1}}, k = 1, 2, \dots, N$ (3.12)

$w_{i,t+\tau}^{\text{name}} \leftarrow w_{i,t+\tau-1}^{\text{name}} \times \frac{c_{i,t+\tau}}{c_{i,t+\tau-1}}, i = 1, 2, \dots, N$ (3.13) $\triangleright t + \tau - 1$ and $t + \tau$ are adjacent intraday timestamps

if $\tau \% \mathcal{T} == 0$ or end of the trading day \triangleright rebalancing point

Calculate cost($t + \tau, w_{(k),t}^{\text{rank}}$) by (3.14)

$V_t \leftarrow V_t - \text{cost}(t + \tau, w_{(k),t}^{\text{rank}})$

$w_{\mathcal{I}_{(k),t+\tau},t+\tau}^{\text{name}} \leftarrow w_{(k),t+\tau}^{\text{rank}}, \quad k = 1, 2, \dots, N$

$\tau \leftarrow \tau + 1$

$V_{t+1} \leftarrow (1 + r_{f,t+1})V_t + \sum_i w_{i,t+\tau}^{\text{name}}$

$w^{\text{prev}} \leftarrow w_{t+\tau}^{\text{name}}$

$t \leftarrow t + 1$

return $V_t, \quad t = 1, 2, \dots, T + 1$

// Input:

// w_t^R : the l_1 -normalized portfolio weights from either parametric
// model (Algorithm 2) or neural networks (Algorithm 3).

// $r_{f,t}$: risk-free rate during the trading day t .

// \mathcal{T} : rebalance interval.

// $c_{t+\tau}$: the capitalization processes in name space and rank space at
// 1-minute resolution throughout the trading day t .
// t is the time tick at daily level.
// τ is the time tick at minute level.

// Output:

// V_t : the value process of the portfolio (PnL) with weights w_t^R .

// Note:
// The algorithm realizes the formulation in section 2.3.1.
// In essence, it converts portfolio weights from rank to name
// at \mathcal{T} minutes interval.

Algorithm 5 portfolio metric [Fig. 1, panel (g)]

Input: V_t , $t = 1, 2, \dots, N$
Output: r_{annual} , σ_{annual} , $\text{SR}_{\text{annual}}$

```

Function portfolio_metric( $V_t, r_f$ )
  For year in years_in_backtesting:
    Locate all  $t_1 \leq t_2 \leq \dots \leq t_N$  in year
     $r_{t_i} = V_{t_i}/V_{t_{i-1}} - 1$ ,  $i = 1, 2, \dots, N$ 
     $r_{\text{annual}} \leftarrow (\prod_{i=1}^N (1 + r_{t_i}))^{252/N} - 1$ 
     $\sigma_{\text{annual}} \leftarrow \sqrt{252} \times \text{std}(\{r_{t_i}\}_{i=1}^N)$   $\triangleright$  std is the standard deviation
     $\text{SR}_{\text{annual}} \leftarrow (r_{\text{annual}} - r_{f,\text{annual}})/\sigma_{\text{annual}}$   $\triangleright r_{f,\text{annual}}$  is annualized risk-free rate
  return  $r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}}$  for all backtesting years

// Input:
//  $V_t$ : the value process (PnL) of the portfolio with weights  $w_t^R$ .
//  $r_{f,t}$ : the risk-free rate at the end of trading day  $t$ .
// Output:
// The algorithm realizes the formulation in section 2.4.
//  $r_{\text{annual}}$ : the annualized return for all backtesting years.
//  $\sigma_{\text{annual}}$ : the annualized volatility for all backtesting years.
//  $\text{SR}_{\text{annual}}$ : the Sharpe ratio for all backtesting years.

```

Algorithm 6 (Integrated) Statistical arbitrage in name space

Input: r_t, r_f, K
Called algorithm: Algorithm 1, Algorithm 2, Algorithm 3, Algorithm 5
Output: $w_t^R, V_t, \text{SR}_{\text{annual}}$

```

Function statistical_arbitrage_in_name_space( $r_t, r_{f,t}, K$ )
   $\epsilon_t, \Phi_t \leftarrow \text{market_decomposition}(r_t, r_{f,t}, K)$  from Algorithm 1
  if parametric model
     $w_t^{R|\text{OU}} \leftarrow \text{portfolio_weights_by_parametric_model}(\epsilon_t, \Phi_t)$  from Algorithm 2
  if neural networks
     $w_t^{R|\text{NN}} \leftarrow \text{portfolio_weights_by_neural_network}(\epsilon_t, \Phi_t)$  from Algorithm 3
  Calculate PnL  $V_t$  by (3.15)
   $r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}} \leftarrow \text{portfolio_metric}(V_t, r_{f,t})$  from Algorithm 5
  return  $w_t^R, V_t, \text{SR}_{\text{annual}}$ 

// Input:
//  $r_t$ : dividend-adjusted daily return in name space.
//  $r_{f,t}$ : risk-free rate at the end of trading day  $t$ .
//  $K$ : number of market factors, predetermined by analyzing
eigenvalue
  // spectrum of the correlation matrix.
// Output:
//  $w_t^R$ : the  $l_1$ -normalized portfolio weights on stock.
//  $V_t$ : the value process (PnL) of the portfolio with weights  $w_t^R$ .
//  $r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}}$ : annualized return, volatility, and Sharpe
  // ratio.

```

Algorithm 7 (Integrated) Statistical arbitrage in rank space

Input: $c_t, c_{t+\tau}, r_f, K$
Called algorithm: Algorithm 1, Algorithm 2, Algorithm 3, Algorithm 4, Algorithm 5
Output: $w_t^R, V_t, r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}}$

```

Function statistical_arbitrage_in_name_space( $c_t, r_{f,t}, K$ )
    Calculate  $\tilde{r}_t$  by (3.5)
     $\tilde{\epsilon}_t, \tilde{\Phi}_t \leftarrow \text{market\_decomposition}(\tilde{r}_t, r_{f,t}, K)$  from Algorithm 1
    if parametric model
         $w_t^{R|\text{OU}} \leftarrow \text{portfolio\_weights\_by\_parametric\_model}(\epsilon_t, \Phi_t)$  from Algorithm 2
    if neural networks
         $w_t^{R|\text{NN}} \leftarrow \text{portfolio\_weights\_by\_neural\_network}(\epsilon_t, \Phi_t)$  from Algorithm 3
     $V_t \leftarrow \text{intraday\_rebalancing}(w_t^{R|\text{NN}}, r_{f,t}, \mathcal{T}, c_{t+\tau})$  from Algorithm 4
     $r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}} \leftarrow \text{portfolio\_metric}(V_t, r_{f,t})$  from Algorithm 5
    return  $w_t^R, V_t, r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}}$ 

// Input:
//  $c_t$ : capitalizations at the end of trading day  $t$ .
//  $c_{t+\tau}$ : capitalization process at 1-minute resolution throughout the
// trading day  $t$ .  $t$  is the time tick at daily level.
//  $\tau$  is the time tick at intraday level.
//  $r_{f,t}$ : risk-free rate at the end of trading day  $t$ .
//  $K$ : number of market factors, predetermined by analyzing
eigenvalue
    // spectrum of the correlation matrix.
// Output:
//  $w_t^R$ : the  $l_1$ -normalized portfolio weights on artificial financial
// instruments that realize  $\tilde{r}_t$ .
//  $V_t$ : the value process (PnL) of the portfolio with weights  $w_t^R$ .
//  $r_{\text{annual}}, \sigma_{\text{annual}}, \text{SR}_{\text{annual}}$ : annualized return, volatility, and Sharpe
// ratio.

```

353 **B Hybrid-Atlas model**

354 In this section, we introduce a hybrid-Atlas model that motivates the definition of return in (3.5) and
 355 hints the qualitative difference between name space and rank space [7, 6, 18].

356 We study an equity market that consists of n stocks with capitalizations $C(t) = (C_1(t), \dots, C_n(t))$,
 357 where $C_i(t)$ represents the capitalization at time t of the asset with name i . We assume that the
 358 log-capitalizations $Y_i(t) := \log C_i(t)$, $i = 1, \dots, n$, satisfy the system of stochastic differential
 359 equations:

$$dY_i(t) = (g_{R_{i,t}} + \gamma_i + \gamma) dt + \sum_{j=1}^n \rho_{i,j} dW_j(t) \\ + \sigma_{R_{i,t}} Y_i(t) dW_i(t), \quad Y_i(0) = y_i, \quad 0 \leq t < \infty \quad (\text{B.1})$$

360 with given initial condition $y = (y_1, \dots, y_n)'$. We assume that $\gamma = 0$ and the system satisfies the
 361 stability condition

$$\sum_{k=1}^n g_k + \sum_{i=1}^n \gamma_i = 0. \quad (\text{B.2})$$

362 Define the log-capitalization process in rank space, $Z_k(t) = Y_{\mathcal{I}_{k,t}}(t)$. Therefore, the rank return
 363 defined in (3.5) can be viewed as the discrete counterpart of $dZ_k(t)$.

364 Theorem B.3 highlights the qualitative difference between residual rank space and name space beyond
 365 linear transformation.

366 **Lemma B.1.** *For continuous semimartingales Y_1, Y_2, \dots, Y_n , the rank process Z_1, Z_2, \dots, Z_n are
 367 continuous semimartingales, and we have*

$$\sum_{k=1}^n L_t(Z_k) = \sum_{i=1}^n L_t(Y_i), \quad \forall t > 0, \quad (\text{B.3})$$

368 where $L_t(Y) = \frac{1}{2\epsilon} \int_0^t \mathbf{1}_{\{-\epsilon < Y_s < \epsilon\}} dY_s$ is the local time accumulated at origin.

369 *Proof.* The proof follows [6] theorem 2.2. □

370 **Lemma B.2.** *For continuous semimartingales Y_1, Y_2, \dots, Y_n and their rank process Z_1, Z_2, \dots, Z_n ,
 371 we have*

$$dZ_k(t) = \sum_{i=1}^n (N_k(t))^{-1} \mathbf{1}_{\{Z_k(t)=X_i(t)\}} dX_i(t) \\ + \sum_{j=k+1}^n (N_k(t))^{-1} dL_t(Z_k - Z_j) - \sum_{j=1}^{k-1} (N_k(t))^{-1} dL_t(Z_j - Z_k). \quad (\text{B.4})$$

372 , where $S_t(k) := \{i : Y_i(t) = Z_k(t)\}$ and $N_k(t)$ is the cardinality of $S_t(k)$.

373 *Proof.* Our proof follows [6] theorem 2.3 closely.

374 Define

$$U = \{u(\cdot) : \mathbb{R}^+ \times \{1, \dots, n\} \rightarrow \{1, \dots, n\}, Z_k(t) = Y_{u_t(k)}(t), \forall t > 0, k = 1, \dots, n\}. \quad (\text{B.5})$$

375 For any $j \in J$, we define u^j of U as

$$u_t^j(k) := \text{the } j_{N_t(k)}\text{-th smallest element of } S_t(k). \quad (\text{B.6})$$

376 In other words, suppose that at time t precisely m of the processes X_1, \dots, X_n have rank k , denoted
 377 as X_{i_1}, \dots, X_{i_m} . Then, $u_t^j(k)$ is the j_m -th smallest among the indices i_1, \dots, i_m .

378 For $u \in U$, we have

$$\begin{aligned}
Z_k(t) - Z_k(0) &= \sum_{i=1}^n \int_0^t \mathbf{1}_{\{u_s(k)=i\}} dY_i(s) + \sum_{i=1}^n \int_0^t \mathbf{1}_{\{u_s(k)=i\}} d(Z_k(s) - Y_i(s)) \\
&= \frac{1}{n!} \sum_{i=1}^n \int_0^t \sum_{j \in J} \mathbf{1}_{\{u_s^j(k)=i\}} dY_i(s) + \frac{1}{n!} \sum_{i=1}^n \int_0^t \sum_{j \in J} \mathbf{1}_{\{u_s^j(k)=i\}} d(X_{(k)}(s) - X_i(s)) \\
&= \sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} dY_i(s) + \sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} d(Z_k(s) - Y_i(s)) \\
&= \sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} dY_i(s) + \sum_{i=1}^n \int_0^t (N_s(k))^{-1} dL_s((Z_k - Y_i)^+) \\
&\quad - \sum_{i=1}^n \int_0^t (N_s(k))^{-1} dL_s((Z_k - Y_i)^-).
\end{aligned} \tag{B.7}$$

379 , where $(\cdot)^+ = \max(\cdot, 0)$ and $(\cdot)^- = \min(\cdot, 0)$. In the second equality, we replace u by u^j for $j \in J$,
380 then summing over all j . In the third equality, we use $\sum_{j \in J} \mathbf{1}_{\{u_s^j(k)=i\}} = \frac{n!}{N_s(k)} \mathbf{1}_{\{X_{(k)}(s)=X_i(s)\}}$.
381 In the fourth equality, we use lemma B.1 and

$$\begin{aligned}
&\sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} d(Z_k(s) - Y_i(s)) \\
&= \sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} d((Z_k(s) - Y_i(s))^+) \\
&\quad - \sum_{i=1}^n \int_0^t (N_s(k))^{-1} \mathbf{1}_{\{Z_k(s)=Y_i(s)\}} d((Z_k(s) - Y_i(s))^-) \\
&= \sum_{i=1}^n \int_0^t (N_s(k))^{-1} dL_s((Z_k - Y_i)^+) - \sum_{i=1}^n \int_0^t (N_s(k))^{-1} dL_s((Z_k - Y_i)^-).
\end{aligned} \tag{B.8}$$

382 \square

383 **Theorem B.3.** *The log-capitalization process in rank space Z_k satisfies*

$$\begin{aligned}
dZ_k(t) &= (g_{\mathcal{R}_{i,t}} + \gamma_i + \gamma) dt + \sum_{j=1}^n \rho_{i,j} dW_j(t) + \sigma_{\mathcal{R}_{i,t}} Y_i(t) dW_i(t), dY_{\mathcal{I}_{k,t}}(t) \\
&\quad + \frac{1}{2}(d\Lambda^{k,k+1}(t) - d\Lambda^{k-1,k}(t))
\end{aligned} \tag{B.9}$$

384 , where $\Lambda^{k,k+1}$ is the local time accumulated at the origin by the non-negative semimartingale
385 $Z_k(t) - Z_{k+1}(t)$ defined as

$$\Lambda^{k,k+1}(t) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbf{1}_{\{-\varepsilon < Z_k(s) - Z_{k+1}(s) < \varepsilon\}} d(Z_k(s) - Z_{k+1}(s)) \tag{B.10}$$

386 *Proof.* From the lemmas above, we have

$$\begin{aligned}
dZ_k(t) &= (g_{\mathcal{R}_{i,t}} + \gamma_i + \gamma) dt + \sum_{j=1}^n \rho_{i,j} dW_j(t) + \sigma_{\mathcal{R}_{i,t}} Y_i(t) dW_i(t), dY_{\mathcal{I}_{k,t}}(t) \\
&\quad + (N_k(t))^{-1} \left[\sum_{\ell=k+1}^n d\Lambda^{k,\ell}(t) - \sum_{\ell=1}^{k-1} d\Lambda^{\ell,k}(t) \right] \\
&= (g_{\mathcal{R}_{i,t}} + \gamma_i + \gamma) dt + \sum_{j=1}^n \rho_{i,j} dW_j(t) + \sigma_{\mathcal{R}_{i,t}} Y_i(t) dW_i(t), dY_{\mathcal{I}_{k,t}}(t) \\
&\quad + \frac{1}{2}(d\Lambda^{k,k+1}(t) - d\Lambda^{k-1,k}(t))
\end{aligned} \tag{B.11}$$

387 , where the second equality follows from [18] lemma 1 that the local times $\Lambda^{k,l}$ generated by triple or
388 higher-order collisions are identically equal to zero. In other words, $\Lambda^{k,l} = 0$ for $|k - l| \geq 2, 1 \leq$
389 $k, l, \leq n$. \square

390 **C Market neutrality**

391 The statistical arbitrage portfolios need to satisfy the market neutrality constraint, $w^T \beta = 0$, so that
 392 the return of the portfolio is independent of market factors

$$w^T(r_t - r_f) = w^T(\beta_t F_t + \epsilon_t) = w^T \epsilon_t. \quad (\text{C.1})$$

393 The following theorem proves that the market neutrality of the portfolio constructed in section 3.1
 394 satisfies the constraints.

395 **Theorem C.1.** *If the portfolio weights satisfies Eq. 3.4 or Eq. 3.8, it is market neutral.*

396 *Proof.* We denote the return matrix $R_t = (r_{t-T+1}, r_{t-T+2}, \dots, r_t) \in \mathbb{R}^{N \times T}$. Assume singular value
 397 decomposition of $R_t - R_f$,

$$R_t - R_f = U \Sigma V^T \quad (\text{C.2})$$

398 , where $R_f \in \mathbb{R}^{1 \times T}$ is the risk-free rate, $U \in \mathbb{R}^{N \times N}$, $\Sigma \in \mathbb{R}^{N \times T}$, and $V^T \in \mathbb{R}^{T \times T}$. Then, the
 399 factors and loadings in Eq. 3.1 and ω_t in Eq. 3.2 becomes

$$F_t = \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_K^T \end{pmatrix}, \quad \beta_t = (u_1, u_2, \dots, u_K) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_K \end{pmatrix}, \quad \omega_t = \begin{pmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_K^{-1} \end{pmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_K^T \end{pmatrix} \quad (\text{C.3})$$

400 , where u_i and v_i are the i -th column of matrix U and V . Then, due to the orthogonality between U
 401 and V ,

$$\Phi_t \beta_t = (I - \beta_t \omega_t) \beta_t = \beta_t - \beta_t (\omega_t \beta_t) = \beta_t - \beta_t = 0, \quad (\text{C.4})$$

402 . Therefore,

$$(w_t^R)^T \beta = (w_t^\epsilon)^T \Phi_t \beta = 0 \quad (\text{C.5})$$

403 \square

404 **D Intraday rebalancing**

405 The portfolio weights calculated in the rank space are assigned to artificial financial instruments
 406 that yield rank returns in continuous-time limits defined in (3.5). To make the constructed portfolio
 407 practically implementable, it is necessary to convert these portfolio weights into stock-based portfolios
 408 in name space.

409 A naive approach is to assign the portfolio weights with correspondence between ranks and names at
 410 the end of each trading day and hold the portfolio throughout the following trading day,

$$w_{i,t} = \sum_{k=1}^N w_{(k),t} \mathbf{1}_{\{\mathcal{R}_{i,t}=k\}}, \quad i = 1, 2, \dots, N. \quad (\text{D.1})$$

411 Unfortunately, this straightforward conversion will not retain the advantages of statistical arbitrage in
 412 rank space, because the performance of the derived portfolio will essentially still depend on returns
 413 in name space rather than the rank returns in continuous-time limit. As indicated by (3.5), the returns
 414 in name and rank space start to diverge in the event of rank switching that frequently occurs for most
 415 ranks at an intraday frequency. It indicates that an effective conversion strategy must appropriately
 416 respond to the rank-switching events.

417 Consequently, we propose an intraday rebalancing mechanism in section 3.3. This mechanism
 418 performs conversion from rank space to name space at a frequency that matches rank-switching
 419 events, even though it results in higher transaction costs due to more frequent trading. In the following,
 420 we carry out an in-depth analysis in a two-stock system to emphasize the crucial role of rank switching
 421 and the rebalancing interval in determining the cost of conversion.

422 **D.1 Portfolio rebalancing through rank switching of two stocks**

423 To elucidate the pivotal role of rank switching in our intraday rebalancing strategy, we examine a
 424 two-stock system depicted in Fig. 8, where the two capitalization processes $c_{t,1}$ and $c_{t,2}$ maintain
 425 their ranks during the rebalancing interval $((k-1)\mathcal{T}, k\mathcal{T}], k \in \mathbb{N}$ and swap their ranks during
 426 $(k\mathcal{T}, (k+1)\mathcal{T}], k \in \mathbb{N}$ (panels (a1-a7)). The red lines in panels (b1-b7) and green lines in panels
 427 (c1-c7) show the dollar portfolio weight on the stock that occupies k -th rank in capitalization,
 428 $w_{(k),\tau}, k = 1, 2$. The orange lines in panels (b1-b7) and blue lines in panels (c1-c7) show the dollar
 429 portfolio weight on the stock that has i -th name index, $w_{i,\tau}, i = 1, 2$. We further calculate and present
 430 in panels (d1-d7) the divergence between the total dollar portfolio weights in rank space and the total
 431 dollar portfolio weights in name space, defined as $w_{(1),t} + w_{(2),t} - w_{1,t} - w_{2,t}$. Panels (e1-e7) shows
 432 the cumulative cost from the bid-ask spread arising from the active trading at the rebalancing point.
 433 We highlight several representative timestamps elaborated below.

434 (i) $t = (k-1)\mathcal{T}^+$ in panels (a1-e1): We invest $w_{(2),t}$ on stock 1 and $w_{(1),t}$ on stock 2 since $c_{1,t} < c_{2,t}$.
 435 Therefore, $w_{(1),t} = w_{2,t}$ and $w_{(2),t} = w_{1,t}$;

436 (ii) $t = k\mathcal{T}$ in panel (a2-e2): The processes evolve towards the rebalancing point $k\mathcal{T}$. The relationship
 437 $w_{(1),t} = w_{2,t}, w_{(2),t} = w_{1,t}$ maintains because there is no rank-swapping between stock 1 and stock
 438 2;

439 (iii) $t = k\mathcal{T}^+$ in panels (a3-e3): At the rebalancing point, no active trading is needed as $w_{(1),t} =$
 440 $w_{2,t}, w_{(2),t} = w_{1,t}$ for $i = 1, 2$, and therefore no divergence (latency cost) or cost from bid-ask
 441 spread incurred;

442 (iv) $k\mathcal{T} < t \leq (k+1)\mathcal{T}$ in panels (a4-e4, a5-e5): The processes evolve towards the rebalancing point
 443 $(k+1)\mathcal{T}$. However, because of the rank switch in capitalization between stock 1 and 2 during the
 444 interval, the dollar-valued portfolio for rank and the dollar-valued portfolio for name start diverging,
 445 i.e. $w_{(1),t} \neq w_{2,t}, w_{(2),t} \neq w_{1,t}$ and reaches a maximum at the next rebalancing point $(k+1)\mathcal{T}$
 446 (panel (d4, d5));

447 (v) $t = (k+1)\mathcal{T}^+$ in panels (a6-e6): We carry out active trading to rebalancing the portfolio
 448 such that $w_{(1),t} = w_{1,t}, w_{(2),t} = w_{2,t}$. This requires cash reserve to compensate (i) divergence
 449 $w_{(1),t} + w_{(2),t} - w_{1,t} - w_{2,t}$ (panel (d6)), and (ii) cost from the bid-ask spread (e6);

450 (vi) $t > (k+1)\mathcal{T}^+$ in panels (a7-e7): the system continues to evolve with $w_{(1),t} = w_{1,t}, w_{(2),t} = w_{2,t}$,
 451 and the divergence becomes zero.

452 From the detailed analysis above, the need for active trading stems from their rank switching during
 453 the balance interval. Furthermore, the latency cost is tied to the divergence of total dollar portfolio
 454 weights between rank space and name space, $w_{(1),t}^{\text{rank}} + w_{(2),t}^{\text{rank}} - w_{1,t}^{\text{name}} - w_{2,t}^{\text{name}}$. This divergence
 455 increases with the interval between the time for rank switching and the time for the subsequent
 456 rebalancing point, suggesting that decreasing rebalancing intervals \mathcal{T} might reduce the risk of large
 457 transaction costs by minimizing latency costs.

458 However, the situation becomes more complex when considering the fluctuating nature of the
 459 capitalization process. In the scenario where two adjacent capitalization processes frequently switch
 460 ranks, as shown in Fig. 9, trading too frequently in response to the instantaneous rank changes can
 461 incur substantial, yet unnecessary costs from bid-ask spread. To illustrate this, we present a similar
 462 two-particle system where fluctuating capitalization processes cross their paths (Fig. 9(a1-a3)). We
 463 calculate dollar portfolio weights in name space and rank space according to the aforementioned
 464 intraday rebalancing strategy, and analyze the divergence of total dollar portfolio weights between rank
 465 space and name space (Fig. 9(b1-b3)), the cumulative latency costs (Fig. 9(c1-c3)), the cumulative
 466 costs from bid-ask spread (Fig. 14(d1-d3)), and the cumulative transaction cost (Fig. 9(e1-e3)). We
 467 consider three scenarios under large, medium, and small rebalancing intervals. Remarkably, our
 468 findings underscore a return-risk trade-off: frequent trading (short rebalancing interval) yields lower
 469 divergence and hence lower risk but incurs higher costs from the bid-ask spread, whereas less frequent
 470 trading (large rebalancing interval) results in higher divergence and risk but lower bid-ask costs.
 471 Thus, selecting an appropriate rebalancing interval is crucial for minimizing overall transaction costs
 472 by balancing between latency costs and costs from bid-ask spread. Indeed, we observe a strong
 473 dependence on the profit and loss (PnL) with different intraday rebalance intervals in our empirical
 474 analysis (Fig. 14).

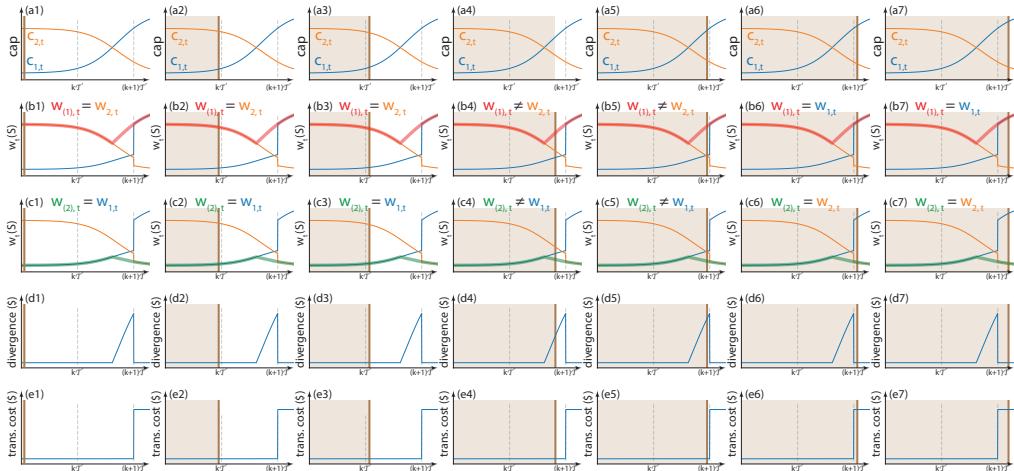


Figure 8: **Schematic for the intraday rebalancing through rank switching of two stocks.** Here, we examine a two-stock system and highlight the critical role of the rank switching. We consider two capitalization processes $c_{t,1}$ and $c_{t,2}$ that maintain their ranks during the rebalancing interval $((k-1)\mathcal{T}, k\mathcal{T}]$ and switch their ranks during $(k\mathcal{T}, (k+1)\mathcal{T}]$ (panel (a1-a7)). The red lines in panel (b1-b7) and green lines in panel (c1-c7) show the dollar-valued portfolio in rank space $w_{(1),\tau}$ and $w_{(2),\tau}$ respectively, where $w_{(k),t}$ denotes the dollar portfolio weights on the stock that occupies k -th rank in capitalization. The orange line in panel (b1-b7) and blue line in panel (c1-c7) show the dollar-valued portfolio in name space $w_{1,\tau}$ and $w_{2,\tau}$ respectively, where $w_{i,\tau}$ denotes the dollar portfolio weights on the stock that had i -th name index. We further calculate and present in panel (d1-d7) the divergence between the total dollar portfolio weights in rank space and in name space, defined as $w_{(1),t} + w_{(2),t} - w_{1,t} - w_{2,t}$. Panel (e1-e7) shows the cumulative cost from bid-ask spread arising from the active trading at the rebalancing point.

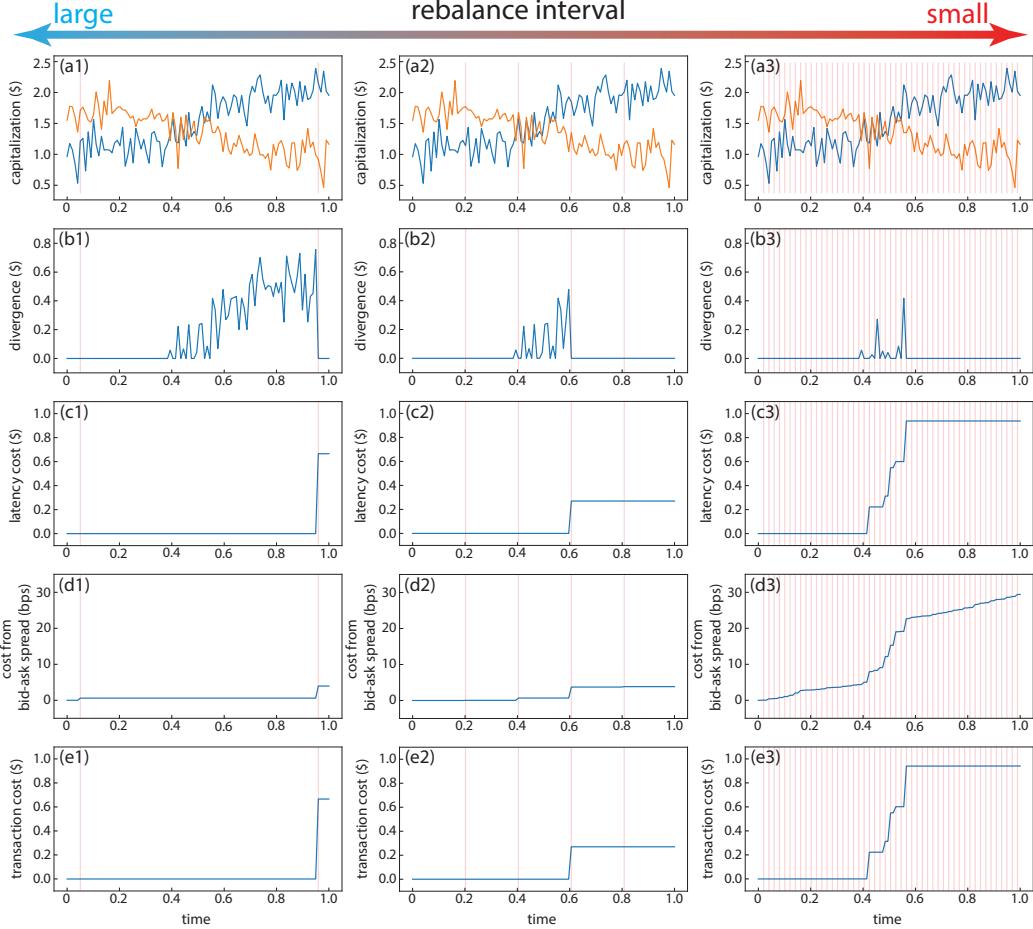


Figure 9: Transaction cost dependence on the rebalancing interval. This figure illustrates how transaction costs incurred by intraday rebalancing are influenced by different rebalancing intervals: large (panel (a1-e1)), medium (panel (a2-e2)), and small (panel (a1-a3)). The red vertical lines mark the rebalancing points across all panels. **(a1-a3)** The capitalization processes from two stocks. **(b1-b3)** The divergence between the total dollar portfolio weights in rank space and name space, $w_{(1),t} + w_{(2),t} - w_{1,t} + w_{2,t}$. The dollar portfolio weights in rank space $w_{(1),t}, w_{(2),t}$, and in name space $w_{1,t}, w_{2,t}$ are calculated based on capitalization processes in (a) following the intraday rebalancing strategy similar to Fig. 8. The maximum divergence decreases as rebalancing interval increases, aiding risk control. **(c1-c3)** The cumulative latency cost required to compensating divergences at the rebalancing points. Each point of rebalancing incurs a latency cost equal to the divergence between the total dollar portfolio weights in rank space and name space. **(d1-d3)** The cumulative cost from the bid-ask spread due to active trading at each rebalancing point. **(e1-e3)** The cumulative transaction cost. The cumulative transaction costs are the sum of latency costs and transaction costs. The medium rebalancing interval results in the lowest transaction costs while maintaining a manageable divergence, illustrating the importance of choosing an appropriate rebalancing interval.

475 **E Benchmark parametric model**

476 Our parametric model serves as the benchmark for DNNs comparison and follows closely to the
 477 framework proposed by Avellaneda and Lee [3] and refined by Yeo and Papanicolaou [28]. This model
 478 applies to both name space and rank space, depending on whether x_t^L or \tilde{x}_t^L is chosen as the input.
 479 We first fit x_t^L to an OU process X_t governed by the stochastic differential equation

$$dX_t = \frac{1}{\tau}(\mu - X_t)dt + \sigma dB_t, \quad (\text{E.1})$$

480 where τ is the mean-reverting time, μ is the long-term average of X_t , B_t is the standard Brownian
 481 motion, and σ is its volatility. Subsequently, we calculate the trading signal in name space

$$s_{i,t}^{\text{OU}} = \frac{x_{i,t} - \hat{\mu}_i}{\hat{\sigma}_i}, \quad (\text{E.2})$$

482 where $\hat{\mu}_i, \hat{\sigma}_i$ are the maximum likelihood estimator of μ and σ [3] and $x_{i,t}$ is the terminal cumulative
 483 residual return at time t ,

$$x_{i,t} = \sum_{j=1}^L \epsilon_{i,t-L+j}. \quad (\text{E.3})$$

484 We also include the estimated mean-reverting time $\hat{\tau}$ to effectively filter the trading opportunities [28].
 485 The details of parameter estimation are presented in the Appendix. We open short/long positions when
 486 observing large positive/negative signals and close positions when the trading signals mean-revert
 487 close to zero (schematic in Fig. 10). Following the principle, the portfolio weights in residual space,
 488 $w_t^{\epsilon|\text{OU,name/rank}}$ become

$$w_{i,t}^{\epsilon|\text{OU,name/rank}} = \begin{cases} -1, & \text{if } w_{i,t-1}^{\epsilon|\text{OU}} = 0, \quad s_{i,t}^{\text{OU}} > c_{\text{thresh-open}}, \quad \hat{\tau}_i < 30 \text{ days} \\ 1, & \text{if } w_{i,t-1}^{\epsilon|\text{OU}} = 0, \quad s_{i,t}^{\text{OU}} < -c_{\text{thresh-open}}, \quad \hat{\tau}_i < 30 \text{ days} \\ 1, & \text{if } w_{i,t-1}^{\epsilon|\text{OU}} = 1, \quad s_{i,t}^{\text{OU}} > c_{\text{thresh-close}}, \quad \hat{\tau}_i < 30 \text{ days} \\ -1, & \text{if } w_{i,t-1}^{\epsilon|\text{OU}} = -1, \quad s_{i,t}^{\text{OU}} > c_{\text{thresh-close}}, \quad \hat{\tau}_i < 30 \text{ days} \\ 0, & \text{otherwise} \end{cases} \quad (\text{E.4})$$

489 For our back-testing, the parameters are set as follows:

$$c_{\text{thresh-open}} = 1.25, \quad c_{\text{thresh-close}} = 0.5, \quad (\text{E.5})$$

490 in accordance with [3, 28]. After calculating $w_t^{\epsilon|\text{OU,name/rank}}$, the conversion to portfolio weights in
 491 equity space, $w_t^{R|\text{OU,name/rank}}$, straightforwardly follow the (3.4) in name space

$$w_t^{R|\text{OU,name}} = \Phi_t^T w_t^{\epsilon|\text{OU,name}} \quad (\text{E.6})$$

492 and from (3.8) in rank space,

$$w_t^{R|\text{OU,rank}} = \tilde{\Phi}_t^T w_t^{\epsilon|\text{OU,name}}. \quad (\text{E.7})$$

493 The practical implementation of the parametric model is summarized in Algorithm 2 along with a
 494 schematic in panel (d1, e1) in Fig. 1.

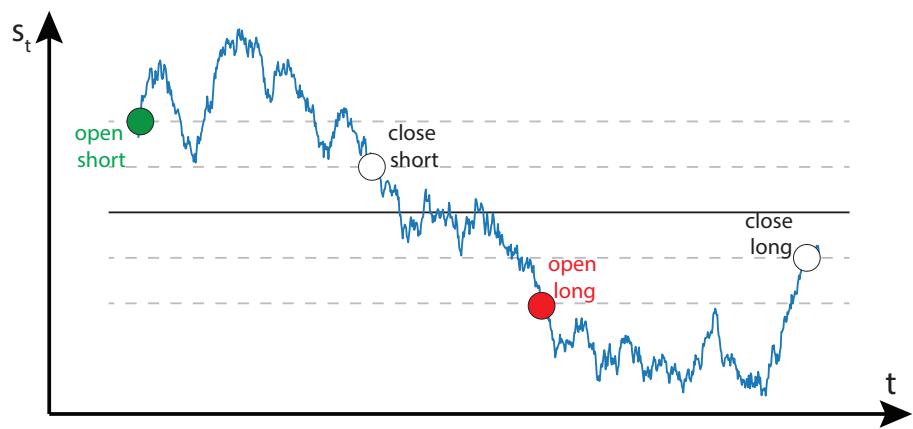


Figure 10: **Schematic for the parametric model.** The blue line shows the trading signal s_t . We open short/long positions when observing large positive/negative signals and close positions when the trading signals mean-revert to zero.

495 **F Non-parametric analysis on mean-reversion of residual returns**

496 The superior mean-reverting behavior in rank space is further demonstrated by comparing the
 497 empirical distribution of normalized cumulative residual returns, \tilde{x}_t^L calculated from name space and
 498 rank space. We define the normalized cumulative residual returns as follows:

$$\hat{x}_t^L = (\hat{x}_{t-L+1}, \hat{x}_{t-L+2}, \dots, \hat{x}_t), \quad \text{where } \hat{x}_{t-L+\alpha} = \frac{1}{\hat{\sigma}_t^L \sqrt{\alpha}} \sum_{j=1}^{\alpha} \epsilon_{t-L+j} \quad (\text{F.1})$$

499 , where $\hat{\sigma}_t^L$ is the estimated standard deviation of $\{\epsilon_{t-L+j}\}_{j=1}^L$. Suppose the residual returns follow
 500 uncorrelated, normal distribution, i.e. $\{\epsilon_{t-L+j}\}_{j=1}^L \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_L)$, the derived cumulative residual
 501 returns x_t^L will follow a standard Brownian motion and the normalized cumulative residual return
 502 defined in (F.1) will be normally distributed, i.e. $\hat{x}_{t-L+\alpha} \sim \mathcal{N}(0, 1)$, $\forall \alpha = 1, 2, \dots, L$. Consequently,
 503 it will serve as a measure of mean-reversion of the difference in probability density function (p.d.f.)
 504 between the empirical observations on market and the normal distribution,

$$\text{p.d.f.}(\hat{x}_{t-L+\alpha}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\hat{x}_{t-L+\alpha}^2}{2}\right), \quad \alpha = 1, 2, \dots, L \quad (\text{F.2})$$

505 The difference is accessed over a series of five-year periods from 1991 to 2022 for both name space
 506 (Fig. 11(a1-a6)) and rank space (Fig. 11(b1-b6)). A more concentrated distribution of $\hat{x}_{t-L+\alpha}$ than
 507 Brownian motion indicates good mean-reverting behavior, especially for large α . This is particularly
 508 evident in rank space, where a robust dominance of red color in the heatmaps for large α regime
 509 (highlighted in dashed boxes in Fig. 11) underscores the concentrated nature of \hat{x}_t^L in rank space.
 510 Such behavior provides critical evidence of a more robust mean-reversion of residual returns in rank
 511 space. In stark contrast, the similar dominance by red color in name space was evident in 1990s
 512 (Fig. 11(a1, a2)), but progressively deteriorated since 2000s (Fig. 11(a5-a6)), and finally disappears
 513 after 2010s (Fig. 11(a5-a6)). This marks the deterioration of the mean-reversion of x_t^L in name space
 514 after the 2010s, echoing the failure of profiting from conventional statistical arbitrage strategies after
 515 2010s.

516 In summary, our non-parametric analysis here demonstrates that the rank space exhibits more robust
 517 mean-reverting behavior compared to name space, echoing the parametric analysis on mean-reverting
 518 time in the main text.

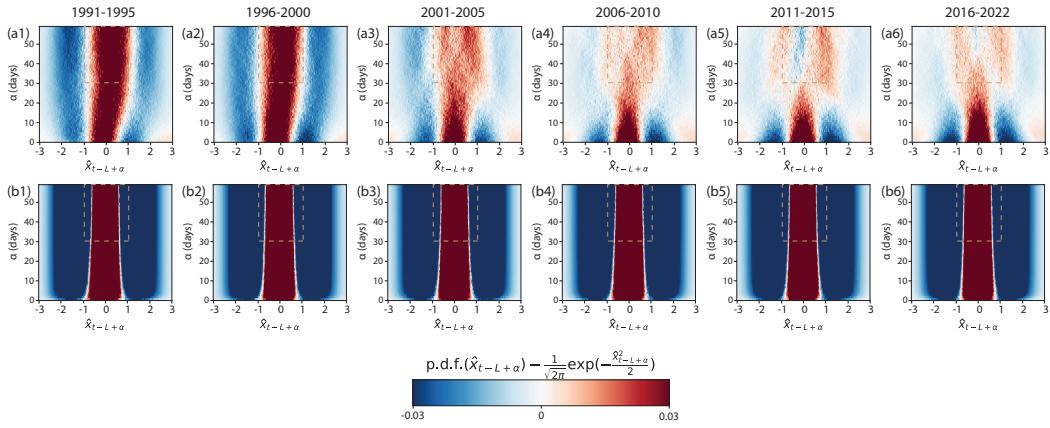


Figure 11: Empirical distributions of normalized cumulative residual returns: name space versus rank space. The cumulative residual returns x_t^L are normalized according to (F.1). This normalization facilitates the conversion of comparisons between trajectories of x_t^L and Brownian motion into comparisons of the probability density of the empirical distribution of \hat{x}_t^L against the normal distribution. Consequently, we present the difference between the empirical probability density of \hat{x}_t^L and standard normal distribution. The empirical probability density of \hat{x}_t^L is evaluated across a series of five-year periods from 1991 (left) to 2022 (right) in both (a1-a6) name space and (b1-b6) rank space. The dashed brown box highlights the critical regime where a dominating red color indicates a more concentrated distribution of $x_{t-L+\alpha}$ for large α . This concentration signals superior mean-reversion capabilities, particularly evident in rank space throughout the last thirty years. Furthermore, the distribution of \hat{x}_t^L in name space evolves significantly, indicating a progressively deteriorating mean-reversion after 2010. This echoes the relatively poor performance in our backtesting. The stark contrast in \hat{x}_t^L supports the strategic advantage of operating in rank space for statistical arbitrage.

519 **G Portfolio performance**

520 **G.1 Annualized return, volatility, and Sharpe ratio**

521 We calculate the annualized return, volatility, and Sharpe ratio derived from the PnL V_t in Fig. 5. The
 522 annualized summary statistics without and with transaction costs are presented in Table 1 and Table 2,
 523 respectively, where we consider corresponding portfolio weights w_t^R calculated by four scenarios: (i)
 524 the parametric benchmark model in name space in panel (a), (ii) the parametric benchmark model in
 525 rank space in panels (b,c), (iii) DNNs in name space in panel (d), and (iv) DNNs in rank space in
 526 panels (e, f).

year	Name space parametric model			rank space parametric model			name space neural networks			rank space neural networks		
	return	vol	SR	return	vol	SR	return	vol	SR	return	vol	SR
2007	2.87%	0.02	1.38	8.52%	0.04	1.90	-8.37%	0.10	-0.87	79.41%	0.14	5.62
2008	7.42%	0.04	1.70	25.44%	0.06	3.96	-6.98%	0.13	-0.52	239.88%	0.38	6.36
2009	7.01%	0.03	2.02	27.81%	0.08	3.57	3.91%	0.13	0.29	414.09%	0.35	11.72
2010	-0.49%	0.02	-0.24	31.73%	0.05	6.21	-0.02%	0.08	0.00	222.37%	0.19	11.41
2011	1.91%	0.02	0.85	40.14%	0.06	7.12	12.84%	0.08	1.67	126.84%	0.22	5.76
2012	-0.40%	0.02	-0.22	41.06%	0.05	8.20	7.35%	0.07	1.00	162.37%	0.20	8.20
2013	1.19%	0.02	0.70	27.92%	0.05	5.74	8.34%	0.07	1.14	289.73%	0.21	13.96
2014	3.65%	0.02	2.07	43.82%	0.05	8.84	-3.24%	0.07	-0.49	168.89%	0.14	12.07
2015	0.81%	0.02	0.41	41.78%	0.06	7.44	0.71%	0.08	0.09	137.95%	0.19	7.10
2016	2.79%	0.02	1.43	61.86%	0.07	9.51	8.58%	0.10	0.90	293.85%	0.27	11.02
2017	1.56%	0.02	0.93	30.58%	0.04	7.04	9.88%	0.07	1.35	208.30%	0.17	12.10
2018	3.07%	0.02	1.44	27.78%	0.05	5.71	3.67%	0.06	0.60	151.91%	0.19	7.83
2019	4.50%	0.02	2.44	41.42%	0.05	8.39	-6.84%	0.06	-1.07	175.91%	0.20	8.59
2020	1.56%	0.04	0.39	25.06%	0.09	2.89	1.24%	0.09	0.14	307.81%	0.28	10.84
2021	-0.67%	0.02	-0.28	37.60%	0.06	6.21	-2.32%	0.08	-0.30	177.60%	0.25	7.11
2022	1.05%	0.03	0.36	36.79%	0.07	5.53	28.84%	0.09	3.29	146.86%	0.29	5.00
Avg	2.36%	0.02	0.96	34.33%	0.06	6.14	3.60%	0.08	0.45	206.49%	0.23	9.04

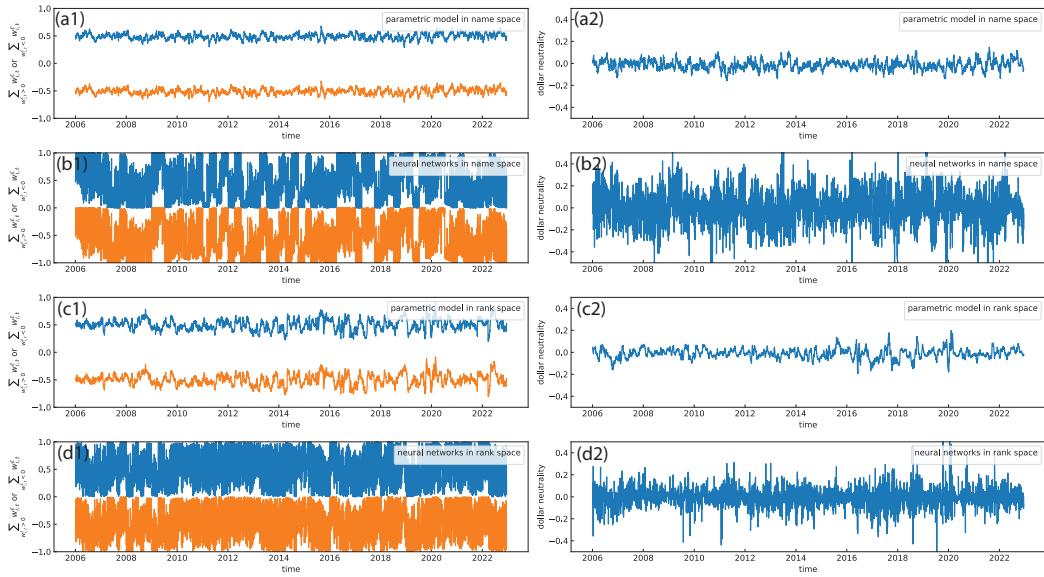
Table 1: **Portfolio performance without transaction costs.** The portfolios in rank space consistently outperform their counterparts in name space, both with the parametric model and neural networks. The neural networks improve the portfolio performance in rank space dramatically, in stark contrast with negligible improvements in name space. The contrast echoes with the fact that the neural networks are much more effective in rank space compared to that in name space Fig. 7.

year	Name space parametric model			rank space parametric model			name space neural networks			rank space neural networks		
	return	vol	SR	return	vol	SR	return	vol	SR	return	vol	SR
2007	1.63%	0.02	0.79	-32.51%	0.04	-7.67	-15.51%	0.10	-1.61	26.07%	0.10	2.55
2008	6.02%	0.04	1.38	-43.78%	0.05	-9.58	-14.19%	0.13	-1.06	36.40%	0.18	2.04
2009	5.71%	0.03	1.65	-32.78%	0.04	-8.07	-3.13%	0.13	-0.23	48.97%	0.13	3.67
2010	-1.69%	0.02	-0.82	-32.83%	0.03	-12.68	-7.25%	0.08	-0.91	43.14%	0.10	4.32
2011	0.71%	0.02	0.31	-32.45%	0.03	-12.02	5.10%	0.08	0.66	14.32%	0.10	1.45
2012	-1.52%	0.02	-0.84	-26.53%	0.02	-11.84	-0.20%	0.07	-0.03	20.41%	0.08	2.42
2013	0.02%	0.02	0.01	-25.14%	0.02	-11.16	0.21%	0.07	0.03	52.51%	0.10	5.37
2014	2.45%	0.02	1.39	-22.52%	0.02	-9.74	-10.25%	0.07	-1.55	35.55%	0.07	4.76
2015	-0.31%	0.02	-0.16	-20.12%	0.03	-7.38	-6.89%	0.08	-0.89	22.82%	0.10	2.32
2016	1.65%	0.02	0.85	-17.68%	0.03	-5.45	0.43%	0.10	0.05	56.09%	0.13	4.31
2017	0.36%	0.02	0.22	-21.38%	0.02	-9.28	2.62%	0.07	0.36	49.00%	0.09	5.16
2018	1.86%	0.02	0.87	-28.83%	0.03	-11.45	-4.16%	0.06	-0.68	27.94%	0.10	2.81
2019	3.34%	0.02	1.81	-21.49%	0.02	-8.92	-13.52%	0.06	-2.12	34.13%	0.10	3.38
2020	0.21%	0.04	0.05	-42.12%	0.05	-9.02	-5.62%	0.09	-0.62	56.62%	0.14	4.14
2021	-2.06%	0.02	-0.86	-36.82%	0.03	-13.27	-9.69%	0.08	-1.27	31.14%	0.13	2.47
2022	-0.16%	0.03	-0.06	-33.01%	0.03	-11.71	19.19%	0.09	2.19	15.69%	0.12	1.26
Avg	1.14%	0.02	0.41	-29.37%	0.03	-9.95	-3.93%	0.08	-0.48	35.68%	0.11	3.28

Table 2: **Portfolio performance with transaction costs.** The portfolio performances in rank space degrade dramatically, for both the parametric model and the neural networks. The substantial degradation arises from the substantial costs associated with realizing rank return in continuous time limit through intraday rebalancing. Nevertheless, the portfolio calculated by neural networks in rank space still yields good results, as the significant transaction costs are compensated by the impressive returns and Sharpe ratio in Table 1, column 4. We choose 2 basis points to account for the transaction costs from bid-ask spread.

527 **G.2 Dollar neutrality**

528 We characterize the long or short proportion of the portfolio weights in equity space, $\sum_{i:w_{i,t}^R > 0} w_{i,t}$
 529 or $\sum_{i:w_{i,t}^R < 0} w_{i,t}$ and the dollar neutrality, $\frac{\sum_i w_{i,t}^R}{\sum_i |w_{i,t}^R|}$. The results are presented in Fig. 12(a1-d1) and
 530 Fig. 12(a2-d2), where we consider w_t^R calculated by four scenarios: (i) the parametric benchmark
 531 model in name space (Fig. 12(a1, a2), (ii) DNNs in name space (Fig. 12(b1, b2), (iii) the parametric
 532 benchmark model in rank space (Fig. 12(c1, c2), (iv) DNNs in rank space (Fig. 12(d1, d2)). Notably,
 533 the long or short proportion of w_t^R by neural networks (Fig. 12(b1, d1)) is much more volatile than
 534 those by the parametric model (Fig. 12(a1, c1)), as a result of flexible leverage adopted by neural
 535 networks. However, the dollar neutrality is satisfied on average thanks to the market neutrality of the
 536 portfolios.



561 **Figure 12: Portfolio weights in residual return and dollar neutrality.** (a1-d1) The temporal
 562 dependence of average long/short portfolio weights in residual space, $\sum_{i:w_{i,t}^e > 0} w_{i,t}^e$ and $\sum_{i:w_{i,t}^e < 0} w_{i,t}^e$.

563 (a2-d2) The deviation from dollar neutrality measured by $\frac{\sum_i w_{i,t}^R}{\sum_i |w_{i,t}^R|}$. We consider four scenarios:
 564 (a1-a2) parametric model in name space; (b1-b2) neural networks in name space; (c1-c2) parametric
 565 model in rank space; (d1-d2) neural networks in rank space.

537 **G.3 Dependence on transaction costs**

538 We present the sensitivity of portfolio performance to transaction costs. We show the PnL with
 539 different transaction cost factor η in Fig. 13(a) and the corresponding Sharpe ratio in Fig. 13. The
 540 current strategy shows significant sensitivity to the transaction cost factor and stops to profit with
 541 $\eta = 5$ basis points. A more effective strategy to realize return in rank space \tilde{r}_t will help the strategy
 542 more immune to transaction costs.

543 **G.4 A characteristic time between rank switching**

544 The rebalancing interval, \mathcal{T} , turns out to be a crucial parameter for statistical arbitrage portfolios
 545 in rank space. To demonstrate, we present the calculated PnL with varying rebalancing intervals
 546 in Fig. 14(a), where the portfolio weights are calculated from the neural networks in rank space.
 547 The associated average Sharpe ratio and terminal PnL as a function of the rebalancing interval are
 548 summarized in Fig. 14(b), with both peaking at 225 minutes.

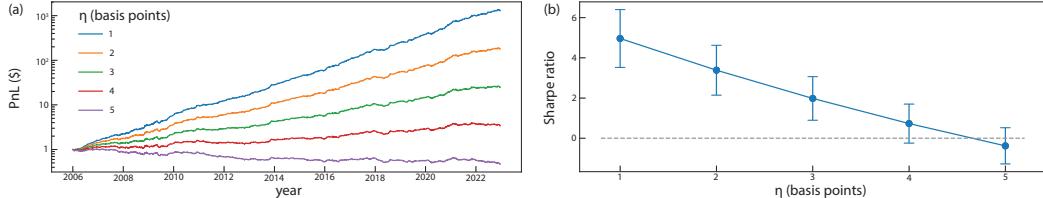


Figure 13: **PnL dependence on transaction cost factor η .** **(a)** The PnL with varying levels of the transaction cost factor η . The underlying portfolio weights are derived from the neural networks in rank space. **(b)** The average Sharpe ratio from 2006 to 2022 at different values of η derived from (a). The strategy stops to profit with 5 basis points transaction costs due to substantial costs associated with realizing rank returns in continuous time limit. The significant change in Sharpe ratio under varying transaction costs underscores the strategy's sensitivity to transaction costs and motivates ongoing improvements in "trading ranks".

549 Here, we delve into the rationale behind the optimal 225-minute interval, starting with an examination
 550 of two proximate capitalization processes modeled as Brownian motions, $c_{1,t}$ and $c_{2,t}$ (Fig. 14(c)).
 551 Given that transaction costs from intraday rebalancing primarily arise from rank swaps in capital-
 552 ization (section 2.3), we measure the cumulative time $\Lambda(t)$ that the capitalization processes cross
 553 (Fig. 14(c), red line),

$$\Lambda(t) = \lim_{\delta \downarrow 0} \int_0^t \mathbf{1}_{\{|c_{1,\tau} - c_{2,\tau}| \leq \delta\}} d\tau \quad (\text{G.1})$$

554 The rank-swapping interval λ is defined as the time between the cross of capitalization processes,
 555 or equivalently, the increments of $\Lambda(t)$. This classical interacting Brownian system features two
 556 characteristic regimes: (i) the idle regime, where the two capitalization processes are distant, main-
 557 taining constant $\Lambda(t)$ with prolonged λ ; (ii) the collision regime (highlighted in brown shaded area in
 558 Fig. 14(c)), where the two capitalization processes stay close, leading to rapid increases in $\Lambda(t)$ and
 559 short λ . We show the empirical distribution of λ on real market in Fig. 14(d), where the small λ values
 560 arise from the collision regime and larger λ values from the idle regime, following approximately
 561 an exponential distribution as a typical signature for standard Brownian particle systems. The 225
 562 minutes is situated at the intersection of the two regimes, establishing it as a characteristic time for
 563 rank switching.

564 In the detailed analysis of the intraday rebalancing in [3.14] and Fig. 9, the transaction costs arise
 565 from (i) latency costs due to delayed reactions post-rank-swapping, and (ii) costs from bid-ask spreads
 566 incurred during active trading. For the collision regime in Fig. 14(c), it is preferable to delay trading
 567 to minimize bid-ask spread costs. Conversely, in the idle regime, immediate trading is preferable to
 568 reduce latency costs. The 225-minute interval effectively differentiates these regimes, thus optimizing
 569 overall transaction costs.

570 The discussion above highlights the challenge in trading ranks – discerning between the collision
 571 and idle regimes and trading at their intersection. Our current intraday rebalancing approach crudely
 572 harnesses average behavior of U.S. equity market, and leaves considerable scope for enhancement
 573 that we will follow up on in subsequent research papers.

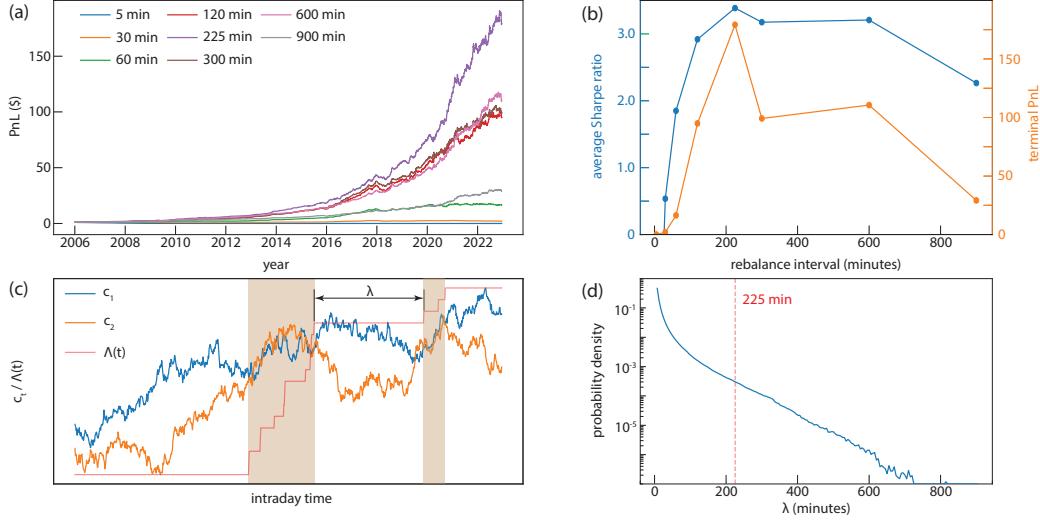


Figure 14: PnL dependence on rebalancing interval and characteristic rank-swapping time. **(a)** The PnL across various intraday rebalancing intervals \mathcal{T} to realize the rank return in continuous time limit \tilde{r}_t . The underlying portfolio weights are derived from the neural networks in rank space. **(b)** Averaged Sharpe ratio between 2006 and 2022 and the terminal PnL as functions of rebalance intervals from (a). Both metrics peak at $\mathcal{T} = 225$ minutes. **(c)** Schematic representation of two capitalization processes, c_1 and c_2 . The $\Lambda(t)$ measures the cumulative time that the capitalization processes c_1 and c_2 cross. Rank switching time λ is the interval between the increments in local time. This stochastic system has two characteristic regimes: (i) the idle regime, where the two capitalization processes are distant, maintaining constant $\Lambda(t)$ with prolonged λ ; (ii) the collision regime (highlighted in brown shaded area in Fig. 14(c)), where the two capitalization processes stay close, leading to rapid increases in $\Lambda(t)$ and short λ . **(e)** The empirical distributions of rank-swapping time τ based on the intraday market data. Low (high) λ arises from the "idle" ("collision") regime. The red dashed line marks the optimal rebalancing interval, $\mathcal{T} = 225$ minutes, positioned at the intersection between the idle and collision regimes, suggesting it is a balanced choice for minimizing transaction costs while responding effectively with rank-swapping events.

574 **References**

- 575 [1] Fernando Acero, Parisa Zehtabi, Nicolas Marchesotti, Michael Cashmore, Daniele Magazzeni,
576 and Manuela Veloso. Deep reinforcement learning and mean-variance strategies for responsible
577 portfolio optimization. *arXiv:2403.16667*, 2024.
- 578 [2] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. arxiv
579 2019. *arXiv preprint arXiv:1908.10063*, 2019.
- 580 [3] Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the US equities market. *Quanti-
581 tative Finance*, 10(7):761–782, 2010.
- 582 [4] Marco Avellaneda, Brian Healy, Andrew Papanicolaou, and George Papanicolaou. Principal
583 eigenportfolios for US equities. *SIAM Journal on Financial Mathematics*, 13(3):702–744, 2022.
- 584 [5] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization.
585 *arXiv:1607.06450*, 2016.
- 586 [6] Adrian D Banner and Raouf Ghomrasni. Local times of ranked continuous semimartingales.
587 *Stochastic Processes and their Applications*, 118(7):1244–1253, 2008.
- 588 [7] Adrian D Banner, Robert Fernholz, and Ioannis Karatzas. Atlas models of equity markets.
589 *Annals of applied probability*, 15(4):2296–2330, 2005.
- 590 [8] Howard C Berg. *Random walks in biology*. Princeton University Press, 1993.
- 591 [9] Jean-Philippe Bouchaud and Marc Potters. Financial applications of random matrix theory: a
592 short review. *arXiv preprint arXiv:0910.1205*, 2009.
- 593 [10] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and high-frequency trading*.
594 Cambridge University Press, 2015.
- 595 [11] E Robert Fernholz and E Robert Fernholz. *Stochastic portfolio theory*. Springer, 2002.
- 596 [12] Cesar Garcia-Diaz. Sociophysics: A physicist’s modeling of psycho-political phenomena
597 (understanding complex systems), 2013.
- 598 [13] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The
599 Review of Financial Studies*, 33(5):2223–2273, 2020.
- 600 [14] Jorge Guijarro-Ordonez, Markus Pelger, and Greg Zanotti. Deep learning statistical arbitrage.
601 *arXiv:2106.04028*, 2021.
- 602 [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image
603 recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*,
604 pages 770–778, 2016.
- 605 [16] Brian Healy, Y.-F. Li, A Papanicolaou, and P Papanicolaou. The data stabilizing transformation
606 from named to ranked equity returns. *working paper*, 2025.
- 607 [17] Blanka Horvath, Aitor Muguruza, and Mehdi Tomas. Deep learning volatility: a deep neural
608 network perspective on pricing and calibration in (rough) volatility models. *Quantitative
609 Finance*, 21(1):11–27, 2021.
- 610 [18] Tomoyuki Ichiba, Vassilios Papathanakos, Adrian Banner, Ioannis Karatzas, and Robert Fern-
611 holz. Hybrid atlas models. *Annals of applied probability*, 21(2):609–644, 2011.
- 612 [19] Taewook Kim and Ha Young Kim. Optimizing the pairs-trading strategy using deep reinforce-
613 ment learning with trading and stop-loss boundaries. *Complexity*, 2019(1):3582516, 2019.
- 614 [20] Tim Siu-tang Leung and Xin Li. *Optimal mean reversion trading: Mathematical analysis and
615 practical applications*, volume 1. World Scientific, 2015.
- 616 [21] Gerald D Mahan. *Many-particle physics*. Springer Science & Business Media, 2013.
- 617 [22] John M Mulvey, Yifan Sun, Mengdi Wang, and Jing Ye. Optimizing a portfolio of mean-
618 reverting assets with transaction costs via a feedforward neural network. *Quantitative Finance*,
619 20(8):1239–1261, 2020.
- 620 [23] Szilárd Pafka and Imre Kondor. Noisy covariance matrices and portfolio optimization ii. *Physica
621 A: Statistical Mechanics and its Applications*, 319:487–494, 2003.
- 622 [24] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks:
623 A deep learning framework for solving forward and inverse problems involving nonlinear partial
624 differential equations. *Journal of Computational physics*, 378:686–707, 2019.

- 625 [25] Agnès Tourin and Raphael Yan. Dynamic pairs trading using the stochastic control approach.
626 *Journal of Economic Dynamics and Control*, 37(10):1972–1981, 2013.
- 627 [26] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing
628 ingredient for fast stylization. *arXiv:1607.08022*, 2016.
- 629 [27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,
630 Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information*
631 *processing systems*, 30, 2017.
- 632 [28] Joongyeub Yeo and George Papanicolaou. Risk control of mean-reversion time in statistical
633 arbitrage. *Risk and Decision Analysis*, 6(4):263–290, 2017.
- 634 [29] Jiongmin Yong and Xun Yu Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*,
635 volume 43. Springer Science & Business Media, 2012.

636 **NeurIPS Paper Checklist**

637 The checklist is designed to encourage best practices for responsible machine learning research,
638 addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove
639 the checklist: **The papers not including the checklist will be desk rejected.** The checklist should
640 follow the references and follow the (optional) supplemental material. The checklist does NOT count
641 towards the page limit.

642 Please read the checklist guidelines carefully for information on how to answer these questions. For
643 each question in the checklist:

- 644 • You should answer [Yes] , [No] , or [NA] .
645 • [NA] means either that the question is Not Applicable for that particular paper or the
646 relevant information is Not Available.
647 • Please provide a short (1–2 sentence) justification right after your answer (even for NA).

648 **The checklist answers are an integral part of your paper submission.** They are visible to the
649 reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it
650 (after eventual revisions) with the final version of your paper, and its final version will be published
651 with the paper.

652 The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation.
653 While "[Yes]" is generally preferable to "[No]", it is perfectly acceptable to answer "[No]" provided a
654 proper justification is given (e.g., "error bars are not reported because it would be too computationally
655 expensive" or "we were unable to find the license for the dataset we used"). In general, answering
656 "[No]" or "[NA]" is not grounds for rejection. While the questions are phrased in a binary way, we
657 acknowledge that the true answer is often more nuanced, so please just use your best judgment and
658 write a justification to elaborate. All supporting evidence can appear either in the main paper or the
659 supplemental material, provided in Appendix. If you answer [Yes] to a question, in the justification
660 please point to the section(s) where related material for the question can be found.

661 **IMPORTANT**, please:

- 662 • **Delete this instruction block, but keep the section heading “NeurIPS paper checklist”**,
663 • **Keep the checklist subsection headings, questions/answers and guidelines below.**
664 • **Do not modify the questions and only use the provided macros for your answers.**

665 **1. Claims**

666 Question: Do the main claims made in the abstract and introduction accurately reflect the
667 paper’s contributions and scope?

668 Answer: [Yes]

669 Justification: The main claims made in the abstract and introduction accurately reflect the
670 paper’s contributions and scope. We claim that (1) operating statistical arbitrage in rank
671 space significantly improves performance over name space when using deep neural networks
672 (DNNs), and (2) this improvement is attributed to the more structured market dynamics
673 and enhanced mean-reversion properties in rank space. These claims are systematically
674 validated through empirical backtesting on U.S. equities from 2007 to 2022, detailed ablation
675 studies, and analysis of the learned neural network behaviors. The paper remains focused
676 on demonstrating these two core contributions without over-claiming broader applicability
677 beyond the presented setup.

678 Guidelines:

- 679 • The answer NA means that the abstract and introduction do not include the claims
680 made in the paper.
681 • The abstract and/or introduction should clearly state the claims made, including the
682 contributions made in the paper and important assumptions and limitations. A No or
683 NA answer to this question will not be perceived well by the reviewers.
684 • The claims made should match theoretical and experimental results, and reflect how
685 much the results can be expected to generalize to other settings.

- 686 • It is fine to include aspirational goals as motivation as long as it is clear that these goals
687 are not attained by the paper.

688 **2. Limitations**

689 Question: Does the paper discuss the limitations of the work performed by the authors?

690 Answer: [Yes]

691 Justification: The paper explicitly discusses the limitations of the work in the *limitations*
692 and *discussion* section. We highlight that the current strategy is sensitive to transaction
693 costs due to frequent intraday rebalancing required to realize rank returns, which may
694 limit its immediate practicality in less liquid markets. We also provide the framework for
695 future optimization of execution strategies to address this limitation. Additionally, while we
696 demonstrate our method in financial markets, we recognize that broader generalization to
697 other many-particle systems is suggested but not empirically validated in this work.

698 Guidelines:

- 699 • The answer NA means that the paper has no limitation while the answer No means that
700 the paper has limitations, but those are not discussed in the paper.
- 701 • The authors are encouraged to create a separate "Limitations" section in their paper.
- 702 • The paper should point out any strong assumptions and how robust the results are to
703 violations of these assumptions (e.g., independence assumptions, noiseless settings,
704 model well-specification, asymptotic approximations only holding locally). The authors
705 should reflect on how these assumptions might be violated in practice and what the
706 implications would be.
- 707 • The authors should reflect on the scope of the claims made, e.g., if the approach was
708 only tested on a few datasets or with a few runs. In general, empirical results often
709 depend on implicit assumptions, which should be articulated.
- 710 • The authors should reflect on the factors that influence the performance of the approach.
711 For example, a facial recognition algorithm may perform poorly when image resolution
712 is low or images are taken in low lighting. Or a speech-to-text system might not be
713 used reliably to provide closed captions for online lectures because it fails to handle
714 technical jargon.
- 715 • The authors should discuss the computational efficiency of the proposed algorithms
716 and how they scale with dataset size.
- 717 • If applicable, the authors should discuss possible limitations of their approach to
718 address problems of privacy and fairness.
- 719 • While the authors might fear that complete honesty about limitations might be used by
720 reviewers as grounds for rejection, a worse outcome might be that reviewers discover
721 limitations that aren't acknowledged in the paper. The authors should use their best
722 judgment and recognize that individual actions in favor of transparency play an impor-
723 tant role in developing norms that preserve the integrity of the community. Reviewers
724 will be specifically instructed to not penalize honesty concerning limitations.

725 **3. Theory Assumptions and Proofs**

726 Question: For each theoretical result, does the paper provide the full set of assumptions and
727 a complete (and correct) proof?

728 Answer: [Yes]

729 Justification: For each theoretical result, we provide the full set of assumptions and complete
730 proofs in the Appendix. In particular, we include (i) a detailed proof of the market neutrality
731 property of the constructed portfolios under the given market decomposition framework,
732 (ii) a hybrid-Atlas model that motivates operating statistical arbitrage in rank space. All
733 mathematical arguments are presented carefully, and the proofs are fully self-contained.

734 Guidelines:

- 735 • The answer NA means that the paper does not include theoretical results.
- 736 • All the theorems, formulas, and proofs in the paper should be numbered and cross-
737 referenced.
- 738 • All assumptions should be clearly stated or referenced in the statement of any theorems.

- 739 • The proofs can either appear in the main paper or the supplemental material, but if
 740 they appear in the supplemental material, the authors are encouraged to provide a short
 741 proof sketch to provide intuition.
 742 • Inversely, any informal proof provided in the core of the paper should be complemented
 743 by formal proofs provided in Appendix or supplemental material.
 744 • Theorems and Lemmas that the proof relies upon should be properly referenced.

745 **4. Experimental Result Reproducibility**

746 Question: Does the paper fully disclose all the information needed to reproduce the main ex-
 747 perimental results of the paper to the extent that it affects the main claims and/or conclusions
 748 of the paper (regardless of whether the code and data are provided or not)?

749 Answer: [Yes]

750 Justification: We provide full disclosure of the information needed to reproduce the main
 751 experimental results. Detailed pseudocode for all key algorithms is included in the Appendix.
 752 In addition, we have uploaded all Python code necessary for calculating the results to a
 753 public GitHub repository. However, the financial market data used in this study cannot be
 754 shared publicly: the daily security data are sourced from CRSP, and the high-frequency
 755 intraday data are sourced from Polygon.io, both of which are under licensing agreements that
 756 prohibit redistribution. Despite this, we provide sufficient algorithmic and implementation
 757 details for independent reproduction using equivalent market datasets.

758 Guidelines:

- 759 • The answer NA means that the paper does not include experiments.
- 760 • If the paper includes experiments, a No answer to this question will not be perceived
 761 well by the reviewers: Making the paper reproducible is important, regardless of
 762 whether the code and data are provided or not.
- 763 • If the contribution is a dataset and/or model, the authors should describe the steps taken
 764 to make their results reproducible or verifiable.
- 765 • Depending on the contribution, reproducibility can be accomplished in various ways.
 766 For example, if the contribution is a novel architecture, describing the architecture fully
 767 might suffice, or if the contribution is a specific model and empirical evaluation, it may
 768 be necessary to either make it possible for others to replicate the model with the same
 769 dataset, or provide access to the model. In general, releasing code and data is often
 770 one good way to accomplish this, but reproducibility can also be provided via detailed
 771 instructions for how to replicate the results, access to a hosted model (e.g., in the case
 772 of a large language model), releasing of a model checkpoint, or other means that are
 773 appropriate to the research performed.
- 774 • While NeurIPS does not require releasing code, the conference does require all submis-
 775 sions to provide some reasonable avenue for reproducibility, which may depend on the
 776 nature of the contribution. For example
 - 777 (a) If the contribution is primarily a new algorithm, the paper should make it clear how
 778 to reproduce that algorithm.
 - 779 (b) If the contribution is primarily a new model architecture, the paper should describe
 780 the architecture clearly and fully.
 - 781 (c) If the contribution is a new model (e.g., a large language model), then there should
 782 either be a way to access this model for reproducing the results or a way to reproduce
 783 the model (e.g., with an open-source dataset or instructions for how to construct
 784 the dataset).
 - 785 (d) We recognize that reproducibility may be tricky in some cases, in which case
 786 authors are welcome to describe the particular way they provide for reproducibility.
 787 In the case of closed-source models, it may be that access to the model is limited in
 788 some way (e.g., to registered users), but it should be possible for other researchers
 789 to have some path to reproducing or verifying the results.

790 **5. Open access to data and code**

791 Question: Does the paper provide open access to the data and code, with sufficient instruc-
 792 tions to faithfully reproduce the main experimental results, as described in supplemental
 793 material?

794 Answer: [Yes]

795 Justification: We provide open access to the full Python code necessary to reproduce the main
796 experimental results on GitHub repository, along with detailed instructions and pseudocode
797 in the supplemental material. However, we cannot provide open access to the data due to
798 licensing restrictions: the daily security data are sourced from CRSP and the high-frequency
799 intraday data from Polygon.io, both of which are proprietary and not publicly distributable.
800 Despite this, all necessary procedures and implementation details are disclosed to enable
801 faithful reproduction using equivalent datasets.

802 Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

822 6. Experimental Setting/Details

823 Question: Does the paper specify all the training and test details (e.g., data splits, hyper-
824 parameters, how they were chosen, type of optimizer, etc.) necessary to understand the
825 results?

826 Answer: [Yes]

827 Justification: We provide all necessary training and testing details. The pseudocode in
828 the Appendix outlines the complete experimental workflow, including data preparation,
829 model training, and evaluation. We also specify the full architecture of the neural network
830 in the appendix, including convolutional and transformer layers, and describe the training
831 procedure, including input features, objective function (mean-variance optimization), and
832 training window sizes. Hyperparameters such as the number of channels, kernel sizes,
833 number of attention heads, and dropout rates are explicitly stated. Our provided pseudocode
834 and implementation details allow faithful reproduction.

835 Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail
that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in Appendix, or as supplemental
material.

841 7. Experiment Statistical Significance

842 Question: Does the paper report error bars suitably and correctly defined or other appropriate
843 information about the statistical significance of the experiments?

844 Answer: [NA]

845 Justification: We report performance using average annual returns and Sharpe ratios across
846 multiple years (2007–2022), with detailed year-by-year statistics provided in the Appendix.

847 Since our experiments are based on real historical market data rather than simulations,
848 traditional error bars or confidence intervals over repeated trials are not applicable. Instead,
849 we present the variability of results across different years to convey the robustness and
850 statistical significance of our findings.

851 Guidelines:

- 852 • The answer NA means that the paper does not include experiments.
- 853 • The authors should answer "Yes" if the results are accompanied by error bars, confi-
854 dence intervals, or statistical significance tests, at least for the experiments that support
855 the main claims of the paper.
- 856 • The factors of variability that the error bars are capturing should be clearly stated (for
857 example, train/test split, initialization, random drawing of some parameter, or overall
858 run with given experimental conditions).
- 859 • The method for calculating the error bars should be explained (closed form formula,
860 call to a library function, bootstrap, etc.)
- 861 • The assumptions made should be given (e.g., Normally distributed errors).
- 862 • It should be clear whether the error bar is the standard deviation or the standard error
863 of the mean.
- 864 • It is OK to report 1-sigma error bars, but one should state it. The authors should
865 preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis
866 of Normality of errors is not verified.
- 867 • For asymmetric distributions, the authors should be careful not to show in tables or
868 figures symmetric error bars that would yield results that are out of range (e.g. negative
869 error rates).
- 870 • If error bars are reported in tables or plots, The authors should explain in the text how
871 they were calculated and reference the corresponding figures or tables in the text.

872 8. Experiments Compute Resources

873 Question: For each experiment, does the paper provide sufficient information on the com-
874 puter resources (type of compute workers, memory, time of execution) needed to reproduce
875 the experiments?

876 Answer: [Yes]

877 Justification: We provide sufficient information on the computational resources used in the
878 Appendix. We specify the hardware configuration, including the type of GPU and CPU,
879 memory capacity, and details of the training process. This information enables readers to
880 estimate the computational requirements needed to reproduce the experiments.

881 Guidelines:

- 882 • The answer NA means that the paper does not include experiments.
- 883 • The paper should indicate the type of compute workers CPU or GPU, internal cluster,
884 or cloud provider, including relevant memory and storage.
- 885 • The paper should provide the amount of compute required for each of the individual
886 experimental runs as well as estimate the total compute.
- 887 • The paper should disclose whether the full research project required more compute
888 than the experiments reported in the paper (e.g., preliminary or failed experiments that
889 didn't make it into the paper).

890 9. Code Of Ethics

891 Question: Does the research conducted in the paper conform, in every respect, with the
892 NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

893 Answer: [Yes]

894 Justification: The research presented in the paper fully conforms to the NeurIPS Code
895 of Ethics. Our work involves only publicly available financial market data (subject to
896 standard licensing restrictions) and does not involve human subjects, private information, or
897 sensitive demographic attributes. We adhere to principles of transparency, reproducibility,
898 and responsible reporting throughout the paper.

899 Guidelines:

- 900 • The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
 901 • If the authors answer No, they should explain the special circumstances that require a
 902 deviation from the Code of Ethics.
 903 • The authors should make sure to preserve anonymity (e.g., if there is a special consid-
 904 eration due to laws or regulations in their jurisdiction).

905 10. Broader Impacts

906 Question: Does the paper discuss both potential positive societal impacts and negative
 907 societal impacts of the work performed?

908 Answer: [Yes]

909 Justification: Our work introduces a rank-based representation of market dynamics that im-
 910 proves the learning efficiency of deep neural networks in noisy environments, with potential
 911 applications beyond finance to complex systems in physics, biology, and social science. By
 912 providing a more structured and stable modeling framework, this approach may facilitate
 913 new insights into the behavior of many-particle systems. However, in financial contexts,
 914 high-frequency statistical arbitrage strategies could contribute to liquidity fragmentation
 915 or increase market volatility if deployed without appropriate safeguards. Our method is
 916 also sensitive to transaction costs, which may limit its accessibility to participants with
 917 unequal technological resources. We encourage future research to explore both beneficial
 918 applications and responsible deployment of rank-based learning methods.

919 Guidelines:

- 920 • The answer NA means that there is no societal impact of the work performed.
 921 • If the authors answer NA or No, they should explain why their work has no societal
 922 impact or why the paper does not address societal impact.
 923 • Examples of negative societal impacts include potential malicious or unintended uses
 924 (e.g., disinformation, generating fake profiles, surveillance), fairness considerations
 925 (e.g., deployment of technologies that could make decisions that unfairly impact specific
 926 groups), privacy considerations, and security considerations.
 927 • The conference expects that many papers will be foundational research and not tied
 928 to particular applications, let alone deployments. However, if there is a direct path to
 929 any negative applications, the authors should point it out. For example, it is legitimate
 930 to point out that an improvement in the quality of generative models could be used to
 931 generate deepfakes for disinformation. On the other hand, it is not needed to point out
 932 that a generic algorithm for optimizing neural networks could enable people to train
 933 models that generate Deepfakes faster.
 934 • The authors should consider possible harms that could arise when the technology is
 935 being used as intended and functioning correctly, harms that could arise when the
 936 technology is being used as intended but gives incorrect results, and harms following
 937 from (intentional or unintentional) misuse of the technology.
 938 • If there are negative societal impacts, the authors could also discuss possible mitigation
 939 strategies (e.g., gated release of models, providing defenses in addition to attacks,
 940 mechanisms for monitoring misuse, mechanisms to monitor how a system learns from
 941 feedback over time, improving the efficiency and accessibility of ML).

942 11. Safeguards

943 Question: Does the paper describe safeguards that have been put in place for responsible
 944 release of data or models that have a high risk for misuse (e.g., pretrained language models,
 945 image generators, or scraped datasets)?

946 Answer: [NA]

947 Justification: Our work does not involve the release of pretrained models, scraped datasets,
 948 or generative models that carry a high risk of misuse. The models we develop are applied
 949 to historical financial market data under standard licensing restrictions, and no sensitive or
 950 private information is used. As such, no additional safeguards beyond responsible reporting
 951 and licensing compliance are necessary.

952 Guidelines:

- 953 • The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external assets used in the paper are properly credited and their licenses respected. We clearly state that the daily financial data are sourced from CRSP and the high-frequency intraday data are obtained from Polygon.io, both under institutional licensing agreements. We also acknowledge that these datasets cannot be publicly redistributed due to licensing restrictions. Any public libraries or frameworks used for model development and training (such as standard deep learning libraries) are cited where appropriate.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New Assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The primary new asset introduced in the paper is the implementation code for our statistical arbitrage framework based on rank-space modeling. We provide detailed pseudocode in the Appendix and release the full Python code with accompanying documentation and instructions on a public GitHub repository. This ensures that others can faithfully reproduce the main results and understand the implementation details.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.

- 1006 • At submission time, remember to anonymize your assets (if applicable). You can either
1007 create an anonymized URL or include an anonymized zip file.

1008 **14. Crowdsourcing and Research with Human Subjects**

1009 Question: For crowdsourcing experiments and research with human subjects, does the paper
1010 include the full text of instructions given to participants and screenshots, if applicable, as
1011 well as details about compensation (if any)?

1012 Answer: [NA]

1013 Justification: Our research does not involve any crowdsourcing experiments, human subjects,
1014 or participant studies. All experiments are conducted on historical financial market data
1015 without human interaction.

1016 Guidelines:

- 1017 • The answer NA means that the paper does not involve crowdsourcing nor research with
1018 human subjects.
1019 • Including this information in the supplemental material is fine, but if the main contribu-
1020 tion of the paper involves human subjects, then as much detail as possible should be
1021 included in the main paper.
1022 • According to the NeurIPS Code of Ethics, workers involved in data collection, curation,
1023 or other labor should be paid at least the minimum wage in the country of the data
1024 collector.

1025 **15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human
1026 Subjects**

1027 Question: Does the paper describe potential risks incurred by study participants, whether
1028 such risks were disclosed to the subjects, and whether Institutional Review Board (IRB)
1029 approvals (or an equivalent approval/review based on the requirements of your country or
1030 institution) were obtained?

1031 Answer: [NA]

1032 Justification: Our research does not involve any study participants or human subjects.
1033 Therefore, no risks to participants exist, no disclosures were necessary, and no Institutional
1034 Review Board (IRB) approvals were required.

1035 Guidelines:

- 1036 • The answer NA means that the paper does not involve crowdsourcing nor research with
1037 human subjects.
1038 • Depending on the country in which research is conducted, IRB approval (or equivalent)
1039 may be required for any human subjects research. If you obtained IRB approval, you
1040 should clearly state this in the paper.
1041 • We recognize that the procedures for this may vary significantly between institutions
1042 and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the
1043 guidelines for their institution.
1044 • For initial submissions, do not include any information that would break anonymity (if
1045 applicable), such as the institution conducting the review.